154679

STIC-Biotech/ChemLib

From: Whiteman, Brian
Sent: Friday, May 27, 2005 9:45 AM
To: STIC-Biotech/ChemLib
Subject: seq search

09/610,313
Barnett et al.

SEQ ID NOs: 30, 31, and 32
1) nucleotide search against databases
2) amino acid search against databases

Thank you,

Brian Whiteman
Remsen, 2D14
mail box 2C18
Patent Examiner - Art Unit 1635
United States Patent and Trademark Office
(571) 272-0764

S

****
STAFF USE ONLY

Searcher:______________
Searcher Phone: 2-
Date Searcher Picked up:____
Date Completed: 6/3
Searcher Prep/Rev. Time:___
Online Time:___________

****
Type of Search

NA#:______ AA#:_______
Interference:____ SPDI:___
S/L:_______ Oligomer:_____
Encode/Transl:___________
Structure#:______ Text:___
Inventor:____ Litigation:___

****
Vendors and cost where applicable
STN:_______________
DIALOG:_______________
QUESTEL/ORBIT:___________
LEXIS/NEXIS:___________
SEQUENCE SYSTEM:________
WWW/Internet:___________
Other(Specify):___________

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: May 31, 2005, 23:20:00 ; Search time 1270.42 Seconds
(without alignments)
11504.726 Million cell updates/sec

Title: US-09-610-313B-30
Perfect score: 2469
Sequence: 1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2469

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2469	100.0	2469	6	ABL39959	Abl39959 Synthetic
2	2469	100.0	2469	12	ADM73764	Adm73764 HIV-1 pol
3	2457	99.5	2457	8	ACA03548	Aca03548 Synthetic
4	2457	99.5	2457	10	ADC13266	Adc13266 DNA of HI
5	2442.2	98.9	2463	6	ABL39960	Abl39960 Synthetic
6	2442.2	98.9	2463	12	ADM73765	Adm73765 HIV-1 pol
7	2434.6	98.6	2457	8	ACA03547	Aca03547 Synthetic
8	2434.6	98.6	2457	10	ADC13265	Adc13265 DNA of HI
9	2415.4	97.8	2457	6	ABL39961	Abl39961 Synthetic
10	2415.4	97.8	2457	12	ADM73766	Adm73766 HIV-1 pol
11	2401.8	97.3	2445	8	ACA03546	Aca03546 Synthetic
12	2401.8	97.3	2445	10	ADC13264	Adc13264 DNA of HI
13	2394.8	97.0	3930	10	ADC13230	Adc13230 DNA of HI
14	2393.2	96.9	3930	10	ADC13231	Adc13231 DNA of HI
15	2393.2	96.9	3930	10	ADC13232	Adc13232 DNA of HI
16	2393.2	96.9	5184	8	ACA03591	Aca03591 Synthetic
17	2393.2	96.9	5184	10	ADC13279	Adc13279 DNA of HI
18	2362.8	95.7	3531	10	ADC13234	Adc13234 DNA of HI
19	2361.2	95.6	3537	10	ADC13236	Adc13236 DNA of HI
20	2360.2	95.6	5145	8	ACA03521	Aca03521 Synthetic
21	2360.2	95.6	5145	10	ADC13233	Adc13233 DNA of HI
22	2350.2	95.2	3538	10	ADC13235	Adc13235 DNA of HI
23	2349.4	95.2	3624	8	ACA03550	Aca03550 Synthetic
24	2349.4	95.2	3624	10	ADC13268	Adc13268 DNA of HI
25	2301.8	93.2	3607	8	ACA03551	Aca03551 Synthetic
26	2301.8	93.2	3607	10	ADC13269	Adc13269 DNA of HI
27	2283.6	92.5	3597	8	ACA03549	Aca03549 Synthetic
28	2283.6	92.5	3597	10	ADC13267	Adc13267 DNA of HI
29	2149.6	87.1	2472	8	ACA03543	Aca03543 Synthetic
30	2149.6	87.1	2472	8	ACC78507	Acc78507 HIV p2Pol
31	2121.2	85.9	2466	8	ACA03542	Aca03542 Synthetic
32	2121.2	85.9	2466	8	ACC78506	Acc78506 HIV p2Pol
33	2094.4	84.8	2460	8	ACA03541	Aca03541 Synthetic
34	2094.4	84.8	2460	8	ACC78505	Acc78505 HIV p2Pol
35	2093.4	84.8	3564	8	ACC78488	Acc78488 HIV GagPo
36	2093.4	84.8	3564	8	ACC78489	Acc78489 HIV GagPo
37	2092.8	84.8	4716	8	ACA03522	Aca03522 Synthetic
38	2092.8	84.8	4716	10	ADC13238	Adc13238 DNA of HI
39	2089	84.6	3999	8	ACC78484	Acc78484 HIV GagCo
40	2087.4	84.5	3999	8	ACC78485	Acc78485 HIV GagCo
41	2087.4	84.5	3999	8	ACC78486	Acc78486 HIV GagCo
42	2087.4	84.5	5283	8	ACA03584	Aca03584 Synthetic
43	2087.4	84.5	5283	8	ACC78529	Acc78529 HIV TatRe
44	2087.2	84.5	4713	8	ACA03592	Aca03592 Synthetic
45	2087.2	84.5	4713	10	ADC13280	Adc13280 DNA of HI

ALIGNMENTS

```
RESULT 1
ABL39959
ID   ABL39959 standard; DNA; 2469 BP.
XX
AC   ABL39959;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 8; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
```

```
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 2469;  DB 6;  Length 2469;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-298;
  Matches 2469;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy        2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy        2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy        2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy        2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy        2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy        2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy        2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy        2461 GGTGAATTC 2469
               |||||||||
Db        2461 GGTGAATTC 2469
```

```
RESULT 2
ADM73764
ID   ADM73764 standard; DNA; 2469 BP.
XX
AC   ADM73764;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #7.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 30; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 2469;  DB 12;  Length 2469;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-298;
  Matches 2469;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy          61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy         121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy         181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy         241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy         301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy         361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy         421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy         481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy         541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy         601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy         661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy         721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
```

```
Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2461 GGTGAATTC 2469
              |||||||||
Db       2461 GGTGAATTC 2469
```

```
RESULT 3
ACA03548
ID   ACA03548 standard; DNA; 2457 BP.
XX
AC   ACA03548;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #31.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
```

```
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 36; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

  Query Match             99.5%;  Score 2457;  DB 8;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-296;
  Matches 2457;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
RESULT 4
ADC13266
ID   ADC13266 standard; DNA; 2457 BP.
XX
AC   ADC13266;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 42; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

  Query Match             99.5%;  Score 2457;  DB 10;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-296;
  Matches 2457;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
RESULT 5
ABL39960
ID   ABL39960 standard; DNA; 2463 BP.
XX
AC   ABL39960;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975YM SEQ ID NO:31.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
```

```
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 9; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

  Query Match             98.9%;  Score 2442.2;  DB 6;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 1.6e-294;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGTGAATTC 2469
              |||||||||
Db       2455 GGTGAATTC 2463


RESULT 6
ADM73765
ID   ADM73765 standard; DNA; 2463 BP.
XX
AC   ADM73765;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #8.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 31; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

  Query Match             98.9%;  Score 2442.2;  DB 12;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 1.6e-294;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
```

```
Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||||      || |  ||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGTGAATTC 2469
              |||||||||
Db       2455 GGTGAATTC 2463

RESULT 7
ACA03547
ID   ACA03547 standard; DNA; 2457 BP.
XX
```

```
AC   ACA03547;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #30.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 35; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match            98.6%;  Score 2434.6;  DB 8;  Length 2457;
  Best Local Similarity  99.6%;  Pred. No. 1.4e-293;
  Matches 2453;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGCTACGAGCTGCACCCC 1374

Qy     1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy     1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy     1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy     1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy     1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy     1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy     1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy     1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy     1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy     1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy     1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy     2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy     2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy     2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy     2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy     2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy     2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy     2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy     2461 GGT 2463
            |||
Db     2455 GGT 2457


RESULT 8
ADC13265
ID   ADC13265 standard; DNA; 2457 BP.
XX
AC   ADC13265;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YM_C SEQ ID NO 44.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 41; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match             98.6%;  Score 2434.6;  DB 10;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 1.4e-293;
  Matches 2453;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy        1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy       61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy      121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy    181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy    241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy    301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy    361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy    421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy    481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy    541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy    601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy    661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy    721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy    781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy    841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy    901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy    961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy   1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy   1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy   1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy   1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
          |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db   1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy   1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy   1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy   1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy   1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy   1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy   1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy   1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy   1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy   1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy   1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy   1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy   1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy   1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy   2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy   2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy   2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy   2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy   2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
```

```
Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGT 2463
              |||
Db       2455 GGT 2457

RESULT 9
ABL39961
ID   ABL39961 standard; DNA; 2457 BP.
XX
AC   ABL39961;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975YMWM SEQ ID NO:32.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 10; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

  Query Match             97.8%;  Score 2415.4;  DB 6;  Length 2457;
  Best Local Similarity   99.3%;  Pred. No. 3.3e-291;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy        1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
               |||||||||||||||||      || |  ||||||||||||||||||||||||||||||
Db        1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy        1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy        1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
               ||||||||||||||||||||||||||||||||||||   ||      |||||||||||||
Db        1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368

Qy        1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428

Qy        1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488

Qy        1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548

Qy        1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608

Qy        1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668

Qy        1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728

Qy        1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788

Qy        1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848

Qy        1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908

Qy        1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968

Qy        1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028

Qy        2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088

Qy        2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148

Qy        2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208

Qy        2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268

Qy        2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328

Qy        2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388

Qy        2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448

Qy        2461 GGTGAATTC 2469
               |||||||||
Db        2449 GGTGAATTC 2457
```

```
RESULT 10
ADM73766
ID   ADM73766 standard; DNA; 2457 BP.
XX
AC   ADM73766;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #9.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 32; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;
```

```
Query Match             97.8%;  Score 2415.4;  DB 12;  Length 2457;
Best Local Similarity   99.3%;  Pred. No. 3.3e-291;
Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||||      || | ||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              |||||||||||||||||||||||||||||||||||   ||      |||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148

Qy        2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208

Qy        2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268

Qy        2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328

Qy        2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388

Qy        2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448

Qy        2461 GGTGAATTC 2469
               |||||||||
Db        2449 GGTGAATTC 2457
```

```
RESULT 11
ACA03546
ID   ACA03546 standard; DNA; 2445 BP.
XX
AC   ACA03546;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #29.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 34; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match             97.3%;  Score 2401.8;  DB 8;  Length 2445;
  Best Local Similarity   99.2%;  Pred. No. 1.6e-289;
  Matches 2438;  Conservative    0;  Mismatches    7;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              |||||||||||      || |  |||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1254

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1314

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              ||||||||||||||||||||||||||||||    ||      ||||||||||||||||||
Db       1315 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1368

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1428

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1488

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1548

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1608

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1668

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1728

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1788

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1848

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1908

Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1968

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2028

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2088

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2089 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2148

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2149 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2208

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2209 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2268

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2269 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2328

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2329 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2388

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2389 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

```
RESULT 12
ADC13264
ID   ADC13264 standard; DNA; 2445 BP.
XX
AC   ADC13264;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YMWM_C SEQ ID NO 43.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 40; 301pp; English.
XX
```

```
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match             97.3%;  Score 2401.8;  DB 10;  Length 2445;
  Best Local Similarity   99.2%;  Pred. No. 1.6e-289;
  Matches 2438;  Conservative    0;  Mismatches    7;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              ||||||||||      || |    ||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1254

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1314

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              ||||||||||||||||||||||||||||||    ||      ||||||||||||||||||
Db       1315 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1368

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1428

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1488

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1548

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1608

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1668

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1728

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1788

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1848

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1908
```

```
Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1968

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2028

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2088

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2089 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2148

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2149 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2208

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2209 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2268

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2269 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2328

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2329 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2388

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2389 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

```
RESULT 13
ADC13230
ID   ADC13230 standard; DNA; 3930 BP.
XX
AC   ADC13230;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmut_C SEQ ID NO 9.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 6; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

  Query Match             97.0%;  Score 2394.8;  DB 10;  Length 3930;
  Best Local Similarity   99.2%;  Pred. No. 1.1e-288;
  Matches 2431;  Conservative    0;  Mismatches    7;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
```

```
Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
              ||||      || |  ||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
              |||||||||||||||||||||||   ||      ||||||||||||||||||||||||||
Db       2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 14
ADC13231
ID   ADC13231 standard; DNA; 3930 BP.
XX
AC   ADC13231;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutAtt_C SEQ ID NO 10.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
```

```
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 7; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;

  Query Match             96.9%;  Score 2393.2;  DB 10;  Length 3930;
  Best Local Similarity   99.2%;  Pred. No. 1.7e-288;
  Matches 2430;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
              ||||      || |  ||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
              |||||||||||||||||||||||      ||       ||||||||||||||||||||||
Db       2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 15
ADC13232
ID   ADC13232 standard; DNA; 3930 BP.
XX
AC   ADC13232;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutIna_C SEQ ID NO 11.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 8; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

  Query Match             96.9%;  Score 2393.2;  DB 10;  Length 3930;
  Best Local Similarity   99.2%;  Pred. No. 1.7e-288;
  Matches 2430;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
              ||||      || |  ||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
              |||||||||||||||||||||||      ||       ||||||||||||||||||||||||
Db       2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

Search completed: June 1, 2005, 11:33:18
Job time : 1280.42 secs

```
OM nucleic - nucleic search, using sw model

Run on:         June  1, 2005, 03:50:55 ; Search time 10564.3 Seconds
                                          (without alignments)
                                          11324.506 Million cell updates/sec

Title:          US-09-610-313B-30
Perfect score:  2469
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2469

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters:      9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the total score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2469	100.0	2469	6	AX455914	AX455914 Sequence
2	2442.2	98.9	2463	6	AX455915	AX455915 Sequence
3	2415.4	97.8	2457	6	AX455916	AX455916 Sequence
4	2052	83.1	2312	6	BD263706	BD263706 Improved
5	2052	83.1	2312	6	CQ870576	CQ870576 Sequence
6	2052	83.1	2312	6	AR373389	AR373389 Sequence
7	2051.2	83.1	9166	6	AX427930	AX427930 Sequence
8	2042.8	82.7	9788	6	AX427936	AX427936 Sequence
9	2028.8	82.2	9169	6	AX427931	AX427931 Sequence
10	2027	82.1	9194	6	AX427925	AX427925 Sequence
11	2027	82.1	12411	6	AX427927	AX427927 Sequence
12	2025.4	82.0	9194	6	AX427926	AX427926 Sequence
13	2025.2	82.0	2306	6	BD263704	BD263704 Improved
14	2025.2	82.0	2306	6	CQ870574	CQ870574 Sequence
15	2025.2	82.0	2306	6	AR373387	AR373387 Sequence
16	2020.4	81.8	9785	6	AX427938	AX427938 Sequence
17	2018.6	81.8	9189	6	AX427921	AX427921 Sequence
18	2017	81.7	9167	6	AX427933	AX427933 Sequence
19	2017	81.7	9170	6	AX427928	AX427928 Sequence
20	2017	81.7	9782	6	AX427935	AX427935 Sequence
21	2017	81.7	9783	6	AX427934	AX427934 Sequence
22	2017	81.7	9792	6	AX427932	AX427932 Sequence
23	2015.6	81.6	4353	12	AF287353	AF287353 Synthetic
24	1999	81.0	3015	6	AX455946	AX455946 Sequence
25	1998.4	80.9	2300	6	BD263705	BD263705 Improved
26	1998.4	80.9	2300	6	CQ870575	CQ870575 Sequence
27	1998.4	80.9	2300	6	AR373388	AR373388 Sequence
28	1982.4	80.3	3009	6	AX455987	AX455987 Sequence
29	1969.4	79.8	4319	6	BD263639	BD263639 Improved
30	1969.4	79.8	4319	6	CQ870498	CQ870498 Sequence
31	1969.4	79.8	4319	6	AR373322	AR373322 Sequence
32	1958.8	79.3	4352	12	AF287352	AF287352 Synthetic
33	1907	77.2	2577	6	AX457088	AX457088 Sequence
34	1905.8	77.2	4343	6	AX306428	AX306428 Sequence
35	1901.8	77.0	4341	6	AX306429	AX306429 Sequence
36	1892.2	76.6	1977	6	AX455954	AX455954 Sequence
37	1879.2	76.1	7897	6	AX427923	AX427923 Sequence
38	1858	75.3	2305	6	BD263702	BD263702 Improved
39	1858	75.3	2305	6	CQ870572	CQ870572 Sequence
40	1858	75.3	2305	6	AR373385	AR373385 Sequence
41	1837	74.4	1965	6	AX455952	AX455952 Sequence
42	1831.2	74.2	2299	6	BD263703	BD263703 Improved
43	1831.2	74.2	2299	6	CQ870573	CQ870573 Sequence
44	1831.2	74.2	2299	6	AR373386	AR373386 Sequence
45	1825.6	73.9	3024	12	AF287355	AF287355 Synthetic

ALIGNMENTS

```
RESULT 1
AX455914
LOCUS       AX455914                2469 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 30 from Patent WO0204493.
ACCESSION   AX455914
VERSION     AX455914.1  GI:21714899
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 30 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1..2469
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975(+)"
ORIGIN

  Query Match             100.0%;  Score 2469;  DB 6;  Length 2469;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-249;
  Matches 2469;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
```

```
Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2461 GGTGAATTC 2469
Db       2461 GGTGAATTC 2469

RESULT 2
AX455915
LOCUS       AX455915                2463 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 31 from Patent WO0204493.
ACCESSION   AX455915
VERSION     AX455915.1  GI:21714900
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 31 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1..2463
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YM"
ORIGIN

  Query Match            98.9%;  Score 2442.2;  DB 6;  Length 2463;
  Best Local Similarity  99.6%;  Pred. No. 1.6e-246;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy        1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy        1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy        1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy        1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy        1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy        1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy        2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy        2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy        2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy        2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy        2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy        2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy        2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy        2461 GGTGAATTC 2469
               |||||||||
Db        2455 GGTGAATTC 2463


RESULT 3
AX455916
LOCUS       AX455916                2457 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 32 from Patent WO0204493.
ACCESSION   AX455916
VERSION     AX455916.1  GI:21714901
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 32 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1..2457
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YMWM"
ORIGIN

  Query Match             97.8%;  Score 2415.4;  DB 6;  Length 2457;
  Best Local Similarity   99.3%;  Pred. No. 1e-243;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy           1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy          61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy         121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy         181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy         241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy         301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy         361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy         421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy         481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy         541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy         601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy         661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy         721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy         781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy         841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy      1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
             |||||||||||||||||       || |  |||||||||||||||||||||||||||||||
Db      1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy      1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy      1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
             ||||||||||||||||||||||||||||||||||    ||      ||||||||||||
Db      1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368

Qy      1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428

Qy      1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488

Qy      1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548

Qy      1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608

Qy      1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668

Qy      1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728

Qy      1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788

Qy      1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848

Qy      1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908

Qy      1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968

Qy      1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028

Qy      2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088

Qy      2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148

Qy      2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208

Qy      2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268

Qy      2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328

Qy      2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388

Qy      2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448

Qy      2461 GGTGAATTC 2469
             |||||||||
Db      2449 GGTGAATTC 2457
```

```
RESULT 4
BD263706
LOCUS       BD263706                2312 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263706
VERSION     BD263706.1  GI:33073474
KEYWORDS    JP 2002533124-A/73.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 73 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/73
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH   Key
                        Location/Qualifiers
            FT   source          1. .2312
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2312
```

```
                /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
ORIGIN

  Query Match            83.1%;  Score 2052;  DB 6;  Length 2312;
  Best Local Similarity  93.6%;  Pred. No. 9.6e-206;
  Matches 2165;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAĜCTGCGCAAGCACC 1297
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
```

```
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317

Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377

Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437

Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469

Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 5
CQ870576
LOCUS       CQ870576                2312 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 84 from Patent EP1433851.
ACCESSION   CQ870576
VERSION     CQ870576.1  GI:52000092
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 84 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..2312
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt(+)"
ORIGIN

  Query Match             83.1%;  Score 2052;  DB 6;  Length 2312;
  Best Local Similarity   93.6%;  Pred. No. 9.6e-206;
  Matches 2165;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCGCAGCGAGG 343

Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757

Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297

Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357

Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417

Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477

Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537

Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597

Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440
```

```
Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 6
AR373389
LOCUS       AR373389                2312 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 84 from patent US 6602705.
ACCESSION   AR373389
VERSION     AR373389.1  GI:40075492
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 84 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1..2312
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             83.1%;  Score 2052;  DB 6;  Length 2312;
  Best Local Similarity   93.6%;  Pred. No. 9.6e-206;
  Matches 2165;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900
```

```
Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312


RESULT 7
AX427930
LOCUS       AX427930                9166 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 168 from Patent WO0232943.
ACCESSION   AX427930
VERSION     AX427930.1  GI:21538017
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 168 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9166
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match            83.1%;  Score 2051.2;  DB 6;  Length 9166;
  Best Local Similarity  92.6%;  Pred. No. 8.2e-206;
  Matches 2192;  Conservative    0;  Mismatches  163;  Indels   13;  Gaps    3;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       3141 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3199

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       3200 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3259

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
```

```
Db     3260 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3319

Qy      365 ---CCCTGAACTTCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421

Db     3320 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3379

Qy      422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481

Db     3380 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3439

Qy      482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541

Db     3440 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3499

Qy      542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601

Db     3500 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3559

Qy      602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661

Db     3560 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3619

Qy      662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721

Db     3620 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3679

Qy      722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781

Db     3680 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3739

Qy      782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841

Db     3740 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3799

Qy      842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901

Db     3800 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3859

Qy      902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961

Db     3860 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3919

Qy      962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021

Db     3920 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3979

Qy     1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081

Db     3980 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4039

Qy     1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141

Db     4040 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4099

Qy     1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201

Db     4100 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4159

Qy     1202 AGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1261

Db     4160 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4219

Qy     1262 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1321

Db     4220 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4279

Qy     1322 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1381

Db     4280 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4339

Qy     1382 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1441

Db     4340 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4399

Qy     1442 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1501

Db     4400 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4459

Qy     1502 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1561

Db     4460 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4519

Qy     1562 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1621

Db     4520 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4579

Qy     1622 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1681

Db     4580 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4639

Qy     1682 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1741

Db     4640 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4699

Qy     1742 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1801

Db     4700 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4759

Qy     1802 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1861

Db     4760 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4819

Qy     1862 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1921

Db     4820 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4879

Qy     1922 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1981

Db     4880 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4939

Qy     1982 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2041

Db     4940 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4999

Qy     2042 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2101

Db     5000 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5059

Qy     2102 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2161

Db     5060 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5119

Qy     2162 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2221

Db     5120 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5179

Qy     2222 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2281

Db     5180 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5239

Qy     2282 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2341

Db     5240 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5299

Qy     2342 GCAAGGTGCTGTTCCTGGACGGCATCGA 2369

Db     5300 GCAAGGTGCTGTTCCTGGACGGCATCGA 5327
```

RESULT 8
AX427936
LOCUS   AX427936   9788 bp   DNA   linear   PAT 20-JUN-2002
DEFINITION Sequence 174 from Patent WO0232943.
ACCESSION   AX427936
VERSION   AX427936.1 GI:21538023

```
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 174 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9788
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.7%;  Score 2042.8;  DB 6;  Length 9788;
  Best Local Similarity   91.8%;  Pred. No. 6.1e-205;
  Matches 2197;  Conservative    0;  Mismatches  177;  Indels   18;  Gaps    3;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       3141 AGGACTGCACCGAGCGACAGGCTAA------TAGGGAAGATCTGGCCTTCCCACAAGGGA 3194

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       3195 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3254

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db       3255 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3314

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db       3315 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3374

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db       3375 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3434

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db       3435 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3494

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db       3495 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3554

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db       3555 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3614

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db       3615 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3674

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db       3675 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3734

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db       3735 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3794

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db       3795 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3854

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db       3855 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3914

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db       3915 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3974

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db       3975 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4034

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db       4035 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4094

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db       4095 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4154

Qy       1202 AGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1261
Db       4155 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4214

Qy       1262 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1321
Db       4215 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4274

Qy       1322 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1381
Db       4275 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4334

Qy       1382 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1441
Db       4335 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4394

Qy       1442 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1501
Db       4395 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4454

Qy       1502 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1561
Db       4455 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4514

Qy       1562 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1621
Db       4515 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4574

Qy       1622 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1681
Db       4575 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4634

Qy       1682 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1741
Db       4635 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4694

Qy       1742 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1801
Db       4695 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4754

Qy       1802 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1861
Db       4755 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4814
```

```
Qy       1862 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1921
Db       4815 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4874

Qy       1922 TGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1981
Db       4875 TGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4934

Qy       1982 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2041
Db       4935 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4994

Qy       2042 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2101
Db       4995 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5054

Qy       2102 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2161
Db       5055 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5114

Qy       2162 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2221
Db       5115 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5174

Qy       2222 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2281
Db       5175 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5234

Qy       2282 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2341
Db       5235 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5294

Qy       2342 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db       5295 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5346
```

```
RESULT 9
AX427931
LOCUS       AX427931               9169 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 169 from Patent WO0232943.
ACCESSION   AX427931
VERSION     AX427931.1  GI:21538018
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 169 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9169
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.2%;  Score 2028.8;  DB 6;  Length 9169;
  Best Local Similarity   91.7%;  Pred. No. 1.8e-203;
  Matches 2194;  Conservative    0;  Mismatches  182;  Indels   16;  Gaps    4;
```

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2967 TGGCCGAGGCCATGAGCCAGGTGCAGCACACCAACATCATGATGCAGCGCGGCAACTTCA 3026

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       3027 AGGGCCAGAAGCGCAT---CAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3083

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       3084 ACTGCCGCGCCCTGCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 3143

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       3144 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3202

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       3203 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3262

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db       3263 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3322

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db       3323 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3382

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db       3383 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3442

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db       3443 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3502

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db       3503 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3562

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db       3563 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3622

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db       3623 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3682

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db       3683 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3742

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db       3743 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3802

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db       3803 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3862

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db       3863 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3922

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db       3923 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3982

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db       3983 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4042

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db       4043 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4102

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db       4103 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4162

Qy       1202 AGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1261
```

```
Db       4163 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4222

Qy       1262 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1321
Db       4223 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4282

Qy       1322 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1381
Db       4283 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4342

Qy       1382 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1441
Db       4343 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4402

Qy       1442 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1501
Db       4403 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4462

Qy       1502 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1561
Db       4463 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4522

Qy       1562 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1621
Db       4523 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4582

Qy       1622 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1681
Db       4583 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4642

Qy       1682 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1741
Db       4643 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4702

Qy       1742 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1801
Db       4703 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4762

Qy       1802 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1861
Db       4763 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4822

Qy       1862 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1921
Db       4823 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4882

Qy       1922 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1981
Db       4883 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4942

Qy       1982 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2041
Db       4943 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 5002

Qy       2042 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2101
Db       5003 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5062

Qy       2102 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2161
Db       5063 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5122

Qy       2162 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2221
Db       5123 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5182

Qy       2222 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2281
Db       5183 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5242

Qy       2282 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2341
Db       5243 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5302

Qy       2342 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db       5303 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5354
```

```
RESULT 10
AX427925
LOCUS       AX427925                9194 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 163 from Patent WO0232943.
ACCESSION   AX427925
VERSION     AX427925.1  GI:21538012
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 163 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9194
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.1%;  Score 2027;  DB 6;  Length 9194;
  Best Local Similarity   91.6%;  Pred. No. 2.8e-203;
  Matches 2194;  Conservative    0;  Mismatches  185;  Indels   16;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAACT 70
Db       2967 TGGCCGAGGCCATGAGCCAGGTGACCAACAGCGCCACCATCATGATGCAGCGCGGCAACT 3026

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db       3027 TCCGCAACCAGCGCAAGATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCC 3086

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       3087 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 3146

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAG 3205

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       3206 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 3265

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db       3266 TTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 3325

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db       3326 CTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3385

Qy        419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db       3386 TAGGGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGG 3445

Qy        479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db       3446 AGATGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 3505

Qy        539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
```

```
Db   3506 AGGTGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCG 3565

Qy    599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
Db   3566 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 3625

Qy    659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
Db   3626 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 3685

Qy    719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
Db   3686 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 3745

Qy    779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
Db   3746 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 3805

Qy    839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
Db   3806 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 3865

Qy    899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
Db   3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 3925

Qy    959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
Db   3926 ACCCCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 3985

Qy   1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
Db   3986 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 4045

Qy   1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
Db   4046 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 4105

Qy   1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db   4106 GCCCCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 4165

Qy   1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258
Db   4166 CCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 4225

Qy   1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318
Db   4226 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 4285

Qy   1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378
Db   4286 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 4345

Qy   1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438
Db   4346 CCGACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 4405

Qy   1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
Db   4406 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 4465

Qy   1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
Db   4466 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGA 4525

Qy   1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
Db   4526 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 4585

Qy   1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
Db   4586 GCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCC 4645

Qy   1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
Db   4646 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 4705

Qy   1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
Db   4706 GCATGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 4765

Qy   1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
Db   4766 CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGG 4825

Qy   1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
Db   4826 AGACCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 4885

Qy   1919 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
Db   4886 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 4945

Qy   1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
Db   4946 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 5005

Qy   2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
Db   5006 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGA 5065

Qy   2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
Db   5066 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 5125

Qy   2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218
Db   5126 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 5185

Qy   2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278
Db   5186 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 5245

Qy   2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338
Db   5246 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCA 5305

Qy   2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db   5306 TCCGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5360
```

RESULT 11
AX427927

```
LOCUS       AX427927             12411 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 165 from Patent WO0232943.
ACCESSION   AX427927
VERSION     AX427927.1  GI:21538014
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 165 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..12411
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN
```

```
Query Match          82.1%;  Score 2027;  DB 6;  Length 12411;
Best Local Similarity   91.6%;  Pred. No. 2.6e-203;
Matches 2194;  Conservative    0;  Mismatches  185;  Indels   16;  Gaps    4;

Qy       14 TGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAACT 70
Db     2967 TGGCCGAGGCCATGAGCCAGGTGACCAACAGCGCCACCATCATGATGCAGCGCGGCAACT 3026

Qy       71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db     3027 TCCGCAACCAGCGCAAGATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCC 3086

Qy      131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db     3087 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 3146

Qy      191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db     3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTTAGGGAAGATCTGGCCTTCCCACAAG 3205

Qy      251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db     3206 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 3265

Qy      311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db     3266 TTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 3325

Qy      365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db     3326 CTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3385

Qy      419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db     3386 TAGGGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGG 3445

Qy      479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db     3446 AGATGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 3505

Qy      539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db     3506 AGGTGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCG 3565

Qy      599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
Db     3566 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 3625

Qy      659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
Db     3626 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 3685

Qy      719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
Db     3686 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 3745

Qy      779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
Db     3746 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 3805

Qy      839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
Db     3806 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 3865

Qy      899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
Db     3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 3925

Qy      959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
Db     3926 ACCCCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 3985

Qy     1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
Db     3986 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 4045

Qy     1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
Db     4046 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 4105

Qy     1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db     4106 GCCCCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 4165

Qy     1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258
Db     4166 CCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 4225

Qy     1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318
Db     4226 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 4285

Qy     1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378
Db     4286 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 4345

Qy     1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438
Db     4346 CCGACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 4405

Qy     1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
Db     4406 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 4465

Qy     1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
Db     4466 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGA 4525

Qy     1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
Db     4526 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 4585

Qy     1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
Db     4586 GCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCC 4645

Qy     1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
Db     4646 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 4705

Qy     1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
Db     4706 GCATGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 4765

Qy     1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
Db     4766 CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGG 4825

Qy     1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
Db     4826 AGACCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 4885

Qy     1919 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
Db     4886 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 4945

Qy     1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
Db     4946 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 5005

Qy     2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
Db     5006 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGA 5065

Qy     2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
```

```
Db       5066 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 5125

Qy       2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218

Db       5126 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 5185

Qy       2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278

Db       5186 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 5245

Qy       2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338

Db       5246 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCA 5305

Qy       2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393

Db       5306 TCCGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5360
```

```
RESULT 12
AX427926
LOCUS       AX427926                9194 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 164 from Patent WO0232943.
ACCESSION   AX427926
VERSION     AX427926.1  GI:21538013
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 164 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9194
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.0%;  Score 2025.4;  DB 6;  Length 9194;
  Best Local Similarity   91.6%;  Pred. No. 4.1e-203;
  Matches 2193;  Conservative    0;  Mismatches  186;  Indels   16;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAACT 70

Db       2967 TGGCCGAGGCCATGAGCCAGGTGACCAACAGCGCCACCATCATGATGCAGCGCGGCAACT 3026

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130

Db       3027 TCCGCAACCAGCGCAAGATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCC 3086

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190

Db       3087 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 3146

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250

Db       3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAG 3205

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310

Db       3206 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 3265

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364

Db       3266 TTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 3325

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
```

```
Db       3326 CTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3385

Qy        419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478

Db       3386 TAGGGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGG 3445

Qy        479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538

Db       3446 AGATGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 3505

Qy        539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598

Db       3506 AGGTGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCG 3565

Qy        599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658

Db       3566 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 3625

Qy        659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718

Db       3626 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 3685

Qy        719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778

Db       3686 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 3745

Qy        779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838

Db       3746 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 3805

Qy        839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898

Db       3806 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 3865

Qy        899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958

Db       3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 3925

Qy        959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018

Db       3926 ACCCCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 3985

Qy       1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078

Db       3986 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 4045

Qy       1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138

Db       4046 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 4105

Qy       1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198

Db       4106 GCCCCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 4165

Qy       1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258

Db       4166 CCGACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCG 4225

Qy       1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318

Db       4226 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 4285

Qy       1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378

Db       4286 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 4345

Qy       1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438

Db       4346 CCGACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 4405

Qy       1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
```

```
Db       4406 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 4465

Qy       1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
Db       4466 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGA 4525

Qy       1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
Db       4526 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 4585

Qy       1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
Db       4586 GCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCC 4645

Qy       1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
Db       4646 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 4705

Qy       1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
Db       4706 GCATGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 4765

Qy       1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
Db       4766 CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGG 4825

Qy       1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
Db       4826 AGACCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 4885

Qy       1919 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
Db       4886 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 4945

Qy       1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
Db       4946 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 5005

Qy       2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
Db       5006 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGA 5065

Qy       2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
Db       5066 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 5125

Qy       2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218
Db       5126 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 5185

Qy       2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278
Db       5186 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 5245

Qy       2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338
Db       5246 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCA 5305

Qy       2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db       5306 TCCGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5360
```

RESULT 13
BD263704

```
LOCUS       BD263704                2306 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263704
VERSION     BD263704.1  GI:33073472
KEYWORDS    JP 2002533124-A/71.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 71 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/71
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH   Key
                      Location/Qualifiers
            FT   source          1. .2306
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match            82.0%;  Score 2025.2;  DB 6;  Length 2306;
  Best Local Similarity  93.3%;  Pred. No. 6.1e-203;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600
```

```
Qy    758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db    601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy    818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db    661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy    878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db    721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy    938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db    781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy    998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db    841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy   1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db    901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy   1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db    961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy   1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db   1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074

Qy   1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db   1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134

Qy   1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db   1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1194

Qy   1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db   1195 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1254

Qy   1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db   1255 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1314

Qy   1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db   1315 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1374

Qy   1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db   1375 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1434

Qy   1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db   1435 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1494

Qy   1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db   1495 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1554

Qy   1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db   1555 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1614

Qy   1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db   1615 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1674

Qy   1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db   1675 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1734

Qy   1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db   1735 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1794

Qy   1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db   1795 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1854

Qy   2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db   1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1914

Qy   2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db   1915 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1974

Qy   2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db   1975 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2034

Qy   2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db   2035 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2094

Qy   2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db   2095 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2154

Qy   2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db   2155 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2214

Qy   2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db   2215 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2274

Qy   2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db   2275 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2306
```

```
RESULT 14
CQ870574
LOCUS       CQ870574                2306 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 82 from Patent EP1433851.
ACCESSION   CQ870574
VERSION     CQ870574.1  GI:52000090
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 82 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..2306
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt.YM"
ORIGIN

  Query Match             82.0%;  Score 2025.2;  DB 6;  Length 2306;
  Best Local Similarity   93.3%;  Pred. No. 6.1e-203;
```

```
Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db       1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1194

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1195 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1254

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1255 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1314

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1315 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1374

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1375 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1434

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1435 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1494

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1495 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1554

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1555 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1614

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1615 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1674

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1675 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1734

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1735 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1794

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1795 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1854

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1914

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1915 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1974

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1975 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2034

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2035 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2094

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2095 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2154
```

```
Qy     2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db     2155 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2214

Qy     2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db     2215 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2274

Qy     2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db     2275 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2306


RESULT 15
AR373387
LOCUS       AR373387                2306 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 82 from patent US 6602705.
ACCESSION   AR373387
VERSION     AR373387.1  GI:40075490
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 82 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1..2306
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             82.0%;  Score 2025.2;  DB 6;  Length 2306;
  Best Local Similarity   93.3%;  Pred. No. 6.1e-203;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy      170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db        1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy      230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db       61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy      290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCGCAGCGAGG 343
Db      121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy      344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db      181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy      398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db      241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy      458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db      301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy      518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db      361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy      578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db      421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy      638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db      481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy      698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db      541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy      758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db      601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy      818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db      661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy      878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db      721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy      938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db      781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy      998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db      841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy     1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db      901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy     1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db      961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy     1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db     1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074

Qy     1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db     1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134

Qy     1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db     1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1194

Qy     1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db     1195 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1254

Qy     1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db     1255 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1314

Qy     1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db     1315 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1374

Qy     1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db     1375 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1434

Qy     1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db     1435 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1494

Qy     1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db     1495 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1554

Qy     1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db     1555 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1614
```

```
Qy   1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
          ||||||||||||||||| |   ||  ||||||||||||||||||||||||| ||||||||
Db   1615 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1674

Qy   1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
          ||   ||||||||||||||||||||||||||| |||||||||  || |||||||||||||
Db   1675 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1734

Qy   1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1735 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1794

Qy   1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
          ||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db   1795 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1854

Qy   2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
          |||||||| | ||||||||||||||||||||||||||||||||||||||| | |||||| 
Db   1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1914

Qy   2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
          |  |||| ||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1915 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1974

Qy   2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
          |||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1975 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2034

Qy   2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   2035 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2094

Qy   2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
          ||||||||||||||  |||||||||||||||||||||||||||||||||||||||| | |
Db   2095 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2154

Qy   2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
          ||||||||||||||   ||||||||||||||||||||||||| |||||||||||||||||
Db   2155 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2214

Qy   2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2215 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2274

Qy   2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
          ||||||||||||||||||||||||||||||||
Db   2275 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2306

Search completed: June  1, 2005, 20:20:25
Job time : 10578.3 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         June  1, 2005, 10:29:57 ; Search time 1414.77 Seconds
                                           (without alignments)
                                           10728.225 Million cell updates/sec

Title:          US-09-610-313B-30
Perfect score:  2469
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2469

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       5706582 seqs, 3073711274 residues

Total number of hits satisfying chosen parameters:      11413164

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                14:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                16:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
                17:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
                18:  /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
                19:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                20:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
                21:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                22:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                       %
Result                Query
  No.      Score      Match Length DB  ID                        Description
-------------------------------------------------------------------------------
    1       2469     100.0    2469  10  US-09-899-575-30          Sequence 30, Appl
    2       2457      99.5    2457  15  US-10-190-435-45          Sequence 45, Appl
    3       2457      99.5    2457  16  US-10-190-305A-39         Sequence 39, Appl
    4     2442.2      98.9    2463  10  US-09-899-575-31          Sequence 31, Appl
    5     2434.6      98.6    2457  15  US-10-190-435-44          Sequence 44, Appl
    6     2434.6      98.6    2457  16  US-10-190-305A-38         Sequence 38, Appl
    7     2415.4      97.8    2457  10  US-09-899-575-32          Sequence 32, Appl
    8     2401.8      97.3    2445  15  US-10-190-435-43          Sequence 43, Appl
    9     2401.8      97.3    2445  16  US-10-190-305A-37         Sequence 37, Appl
   10     2394.8      97.0    3930  15  US-10-190-435-9           Sequence 9, Appli
   11     2393.2      96.9    3930  15  US-10-190-435-10          Sequence 10, Appl
   12     2393.2      96.9    3930  15  US-10-190-435-11          Sequence 11, Appl
   13     2393.2      96.9    5184  15  US-10-190-435-58          Sequence 58, Appl
   14     2393.2      96.9    5184  16  US-10-190-305A-82         Sequence 82, Appl
   15     2362.8      95.7    3531  15  US-10-190-435-13          Sequence 13, Appl
   16     2361.2      95.6    3537  15  US-10-190-435-14          Sequence 14, Appl
   17     2361.2      95.6    3537  15  US-10-190-435-15          Sequence 15, Appl
   18     2360.2      95.6    5145  15  US-10-190-435-12          Sequence 12, Appl
   19     2360.2      95.6    5145  16  US-10-190-305A-12         Sequence 12, Appl
   20     2349.4      95.2    3624  15  US-10-190-435-47          Sequence 47, Appl
   21     2349.4      95.2    3624  16  US-10-190-305A-41         Sequence 41, Appl
   22     2301.8      93.2    3607  15  US-10-190-435-48          Sequence 48, Appl
   23     2301.8      93.2    3607  16  US-10-190-305A-42         Sequence 42, Appl
   24     2283.6      92.5    3597  15  US-10-190-435-46          Sequence 46, Appl
   25     2283.6      92.5    3597  16  US-10-190-305A-40         Sequence 40, Appl
   26     2149.6      87.1    2472  16  US-10-241-009-32          Sequence 32, Appl
   27     2149.6      87.1    2472  16  US-10-190-434B-32         Sequence 32, Appl
   28     2149.6      87.1    2472  16  US-10-190-305A-34         Sequence 34, Appl
   29     2121.2      85.9    2466  16  US-10-241-009-31          Sequence 31, Appl
   30     2121.2      85.9    2466  16  US-10-190-434B-31         Sequence 31, Appl
   31     2121.2      85.9    2466  16  US-10-190-305A-33         Sequence 33, Appl
   32     2094.4      84.8    2460  16  US-10-241-009-30          Sequence 30, Appl
   33     2094.4      84.8    2460  16  US-10-190-434B-30         Sequence 30, Appl
   34     2094.4      84.8    2460  16  US-10-190-305A-32         Sequence 32, Appl
   35     2093.4      84.8    3564  16  US-10-241-009-13          Sequence 13, Appl
   36     2093.4      84.8    3564  16  US-10-241-009-14          Sequence 14, Appl
   37     2093.4      84.8    3564  16  US-10-190-434B-13         Sequence 13, Appl
   38     2093.4      84.8    3564  16  US-10-190-434B-14         Sequence 14, Appl
   39     2092.8      84.8    4716  15  US-10-190-435-17          Sequence 17, Appl
   40     2092.8      84.8    4716  16  US-10-190-305A-13         Sequence 13, Appl
   41       2089      84.6    3999  16  US-10-241-009-9           Sequence 9, Appli
   42       2089      84.6    3999  16  US-10-190-434B-9          Sequence 9, Appli
   43     2087.4      84.5    3999  16  US-10-241-009-10          Sequence 10, Appl
   44     2087.4      84.5    3999  16  US-10-241-009-11          Sequence 11, Appl
   45     2087.4      84.5    3999  16  US-10-190-434B-10         Sequence 10, Appl

                                  ALIGNMENTS

RESULT 1
US-09-899-575-30
; Sequence 30, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
;  APPLICANT: Zur Megede, Jan
;  APPLICANT:  Barnett, Susan W.
;  APPLICANT:  Egnelbrecht, Susan
;  APPLICANT:  van Rensburg, Estrelita Janse
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP01631.102
;  CURRENT APPLICATION NUMBER: US/09/899,575
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: 09/475,704
;  PRIOR FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 135
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 2469
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: PR975(+)
US-09-899-575-30

  Query Match             100.0%;  Score 2469;  DB 10;  Length 2469;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2469;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
```

```
Qy       61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy      121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy      181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy      241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy      301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy      361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy      421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy      481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy      541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy      601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy      661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy      721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy      781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy      841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy      901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy      961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy     1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy     1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy     1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy     1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy     1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy     1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy     1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy     1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy     1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy     1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy     1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy     1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy     1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy     1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy     1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy     1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy     1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy     2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy     2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy     2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy     2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2461 GGTGAATTC 2469
              |||||||||
Db       2461 GGTGAATTC 2469


RESULT 2
US-10-190-435-45
; Sequence 45, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-435-45

  Query Match             99.5%;  Score 2457;  DB 15;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2457;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy     1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy     1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy     1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy     1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy     1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy     1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy     1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy     1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy     1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy     2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy     2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy     2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy     2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy     2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy     2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy     2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 3
US-10-190-305A-39
; Sequence 39, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-305A-39

```
  Query Match             99.5%;  Score 2457;  DB 16;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2457;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy       67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy      127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy      187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy      247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy      307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy      367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy      427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy      487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy      547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy      607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy      667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy      727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780
```

```
Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 4
US-09-899-575-31
; Sequence 31, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
;  APPLICANT: Zur Megede, Jan
;  APPLICANT:  Barnett, Susan W.
;  APPLICANT:  Egnelbrecht, Susan
;  APPLICANT:  van Rensburg, Estrelita Janse
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP01631.102
;  CURRENT APPLICATION NUMBER: US/09/899,575
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: 09/475,704
;  PRIOR FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 135
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 2463
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: PR975YM
US-09-899-575-31

  Query Match             98.9%;  Score 2442.2;  DB 10;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
```

```
Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy         2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy         2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy         2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy         2461 GGTGAATTC 2469
                |||||||||
Db         2455 GGTGAATTC 2463
```

```
RESULT 5
US-10-190-435-44
; Sequence 44, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-435-44

  Query Match             98.6%;  Score 2434.6;  DB 15;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2453;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy            1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy           61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy          121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy          181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy          241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy          301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy          361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy          421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
                |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy          481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy          541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy          601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy          661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy          721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy          781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy          841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy          901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy          961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy         1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy         1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy         1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy         1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
                |||||||||||||||||      || |  ||||||||||||||||||||||||||||||
Db         1201 GAGATCGTGATCTACCA------GGCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy         1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy         1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy         1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy         1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGT 2463
              |||
Db       2455 GGT 2457
```

RESULT 6

```
US-10-190-305A-38
; Sequence 38, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-305A-38

  Query Match             98.6%;  Score 2434.6;  DB 16;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2453;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
```

```
Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGT 2463
              |||
Db       2455 GGT 2457
```

```
RESULT 7
US-09-899-575-32
; Sequence 32, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
;  APPLICANT: Zur Megede, Jan
;  APPLICANT:  Barnett, Susan W.
;  APPLICANT:  Egnelbrecht, Susan
;  APPLICANT:  van Rensburg, Estrelita Janse
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP01631.102
;  CURRENT APPLICATION NUMBER: US/09/899,575
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: 09/475,704
;  PRIOR FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 135
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: PR975YMWM
```

```
US-09-899-575-32

  Query Match             97.8%;  Score 2415.4;  DB 10;  Length 2457;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATÇAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||   ||      |||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
```

```
Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Db       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Db       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Db       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Db       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Db       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Db       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448

Qy       2461 GGTGAATTC 2469
Db       2449 GGTGAATTC 2457
```

```
RESULT 8
US-10-190-435-43
; Sequence 43, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;   LENGTH: 2445
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMWM_C
US-10-190-435-43

  Query Match             97.3%;  Score 2401.8;  DB 15;  Length 2445;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 2438;  Conservative    0;  Mismatches    7;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
Db       1201 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1254

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
Db       1255 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1314
```

```
Qy       1327 AAGAAGCACCAGAAGGAGCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              |||||||||||||||||||||||||||||    ||      |||||||||||||||||
Db       1315 AAGAAGCACCAGAAGGAGCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1368

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1428

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1488

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1548

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1608

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1668

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1728

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1788

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1848

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1908

Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1968

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2028

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2088

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2089 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2148

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2149 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2208

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2209 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2268

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2269 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2328

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2329 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2388

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2389 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445

RESULT 9
US-10-190-305A-37
; Sequence 37, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
;   LENGTH: 2445
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: p2Pol.opt.YMWM_C
US-10-190-305A-37

  Query Match             97.3%;  Score 2401.8;  DB 16;  Length 2445;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 2438;  Conservative    0;  Mismatches    7;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              |||||||||||      || |  |||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1254

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1314

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              ||||||||||||||||||||||||||||||   ||      |||||||||||||||||||
Db       1315 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1368

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1428

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1488

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1548

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1608

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1668

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1728

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1788

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1848

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1908

Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1968

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2028

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2088

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2089 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2148

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2149 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2208

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2209 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2268

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2269 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2328

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2329 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2388

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2389 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 10
US-10-190-435-9
; Sequence 9, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:

```
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmut_C
US-10-190-435-9

  Query Match             97.0%;  Score 2394.8;  DB 15;  Length 3930;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 2431;  Conservative    0;  Mismatches    7;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
              ||||      || |  ||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
              |||||||||||||||||||||||   ||      ||||||||||||||||||||||||||
Db       2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574
```

```
Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 11
US-10-190-435-10
; Sequence 10, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt_C
US-10-190-435-10

  Query Match             96.9%;  Score 2393.2;  DB 15;  Length 3930;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 2430;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
              ||||      || | |||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
              |||||||||||||||||||||      ||      |||||||||||||||||||||||||
Db       2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854
```

```
Qy     1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy     1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy     1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy     1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy     1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy     1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy     1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy     1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy     1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy     1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy     1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy     2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy     2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy     2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy     2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy     2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy     2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy     2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 12
US-10-190-435-11
; Sequence 11, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna_C
US-10-190-435-11

  Query Match             96.9%;  Score 2393.2;  DB 15;  Length 3930;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 2430;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy       14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy       74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy      134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy      194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy      254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy      314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy      374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy      434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
            ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy      494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy      554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy      614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146
```

```
Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
              ||||      || |  ||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
              |||||||||||||||||||||||   ||      ||||||||||||||||||||||||||
Db       2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 13
US-10-190-435-58
; Sequence 58, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
;   LENGTH: 5184
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: TatRevNefgagCpolIna C
US-10-190-435-58
```

```
Query Match             96.9%;  Score 2393.2;  DB 15;  Length 5184;
Best Local Similarity   99.2%;  Pred. No. 0;
Matches 2430;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy       1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
Db       3941 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 3994

Qy       1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
Db       3995 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 4054

Qy       1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
Db       4055 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 4108

Qy       1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
Db       4109 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 4168

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
Db       4169 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 4228

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
Db       4229 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 4288

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
Db       4289 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 4348

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
Db       4349 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 4408

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
Db       4409 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 4468

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
Db       4469 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 4528

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
Db       4529 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 4588

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
Db       4589 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 4648

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
Db       4649 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 4708

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
Db       4709 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 4768

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
Db       4769 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 4828

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4829 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 4888

Qy        2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4889 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 4948

Qy        2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4949 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 5008

Qy        2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5009 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 5068

Qy        2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5069 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 5128

Qy        2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5129 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 14
US-10-190-305A-82
; Sequence 82, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
;   LENGTH: 5184
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: TatRevNefgagCpolIna C
US-10-190-305A-82

```
  Query Match             96.9%;  Score 2393.2;  DB 16;  Length 5184;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 2430;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy          14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy          74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy         134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy         194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy         254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy         314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy         374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy         434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
               ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy         494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy         554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy         614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy         674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy         734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy         794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy         854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy         914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy         974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy        1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy        1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy        1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy        1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
               ||||      || |  ||||||||||||||||||||||||||||||||||||||||||||
Db        3941 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 3994

Qy        1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3995 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 4054

Qy        1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
               |||||||||||||||||||||||   ||      ||||||||||||||||||||||||||
Db        4055 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 4108

Qy        1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4109 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 4168

Qy       1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4169 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 4228

Qy       1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4229 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 4288

Qy       1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4289 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 4348

Qy       1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4349 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 4408

Qy       1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4409 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 4468

Qy       1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4469 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 4528

Qy       1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4529 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 4588

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4589 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 4648

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4649 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 4708

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4709 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 4768

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4769 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 4828

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4829 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 4888

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4889 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 4948

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4949 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 5008

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5009 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 5068

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5069 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 5128

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5129 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 15
US-10-190-435-13

```
; Sequence 13, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 3531
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_C
US-10-190-435-13

  Query Match             95.7%;  Score 2362.8;  DB 15;  Length 3531;
  Best Local Similarity   98.4%;  Pred. No. 0;
  Matches 2411;  Conservative    0;  Mismatches   27;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              ||||||||| ||||||||||||  || |  |   ||  ||  | || ||| |    |
Db       1088 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1147

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1207

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1267

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1327

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1387

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1447

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1507

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1508 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1567

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1568 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1627

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1628 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1687

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1688 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1747

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        1748 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1807

Qy         734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1808 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1867

Qy         794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1868 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1927

Qy         854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1928 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987

Qy         914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2047

Qy         974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2048 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2107

Qy        1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2108 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2167

Qy        1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2168 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2227

Qy        1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2228 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2287

Qy        1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
               ||||      || |  ||||||||||||||||||||||||||||||||||||||||||||
Db        2288 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2341

Qy        1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2342 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2401

Qy        1334 ACCAGAAGGAGCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
               ||||||||||||||||||||||   ||       |||||||||||||||||||||||||||
Db        2402 ACCAGAAGGAGCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2455

Qy        1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2456 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2515

Qy        1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2516 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2575

Qy        1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2576 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 2635

Qy        1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2636 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 2695

Qy        1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2696 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 2755

Qy        1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2756 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 2815

Qy        1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2816 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 2875

Qy        1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2876 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 2935

Qy        1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2936 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 2995

Qy        1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2996 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3055

Qy        1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3056 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3115

Qy        2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3116 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3175

Qy        2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3176 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3235

Qy        2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3236 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3295

Qy        2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3296 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3355

Qy        2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3356 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3415

Qy        2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3416 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3475

Qy        2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3476 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3525
```

Search completed: June 2, 2005, 04:23:32
Job time : 1424.77 secs

```
OM nucleic - nucleic search, using sw model

Run on:          June  1, 2005, 10:29:57 ; Search time 1411.33 Seconds
                                            (without alignments)
                                            10728.225 Million cell updates/sec

Title:           US-09-610-313B-31
Perfect score:   2463
Sequence:        1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2463

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        5706582 seqs, 3073711274 residues

Total number of hits satisfying chosen parameters:      11413164

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                 1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                 2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                 3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                 4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                 5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                 6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                 7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                 9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                 10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                 11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                 12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                 13:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                 14:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                 15:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                 16:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
                 17:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
                 18:  /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
                 19:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                 20:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
                 21:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                 22:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2463	100.0	2463	10	US-09-899-575-31	Sequence 31, Appl
2	2455.4	99.7	2457	15	US-10-190-435-44	Sequence 44, Appl
3	2455.4	99.7	2457	16	US-10-190-305A-38	Sequence 38, Appl
4	2442.2	99.2	2469	10	US-09-899-575-30	Sequence 30, Appl
5	2436.2	98.9	2457	10	US-09-899-575-32	Sequence 32, Appl
6	2430.2	98.7	2457	15	US-10-190-435-45	Sequence 45, Appl
7	2430.2	98.7	2457	16	US-10-190-305A-39	Sequence 39, Appl
8	2422.6	98.4	2445	15	US-10-190-435-43	Sequence 43, Appl
9	2422.6	98.4	2445	16	US-10-190-305A-37	Sequence 37, Appl
10	2415.6	98.1	3930	15	US-10-190-435-9	Sequence 9, Appli
11	2414	98.0	3930	15	US-10-190-435-10	Sequence 10, Appl
12	2414	98.0	3930	15	US-10-190-435-11	Sequence 11, Appl
13	2414	98.0	5184	15	US-10-190-435-58	Sequence 58, Appl
14	2414	98.0	5184	16	US-10-190-305A-82	Sequence 82, Appl
15	2383.6	96.8	3531	15	US-10-190-435-13	Sequence 13, Appl
16	2382	96.7	3537	15	US-10-190-435-14	Sequence 14, Appl
17	2382	96.7	3537	15	US-10-190-435-15	Sequence 15, Appl
18	2381	96.7	5145	15	US-10-190-435-12	Sequence 12, Appl
19	2381	96.7	5145	16	US-10-190-305A-12	Sequence 12, Appl
20	2322.6	94.3	3607	15	US-10-190-435-48	Sequence 48, Appl
21	2322.6	94.3	3607	16	US-10-190-305A-42	Sequence 42, Appl
22	2322.6	94.3	3624	15	US-10-190-435-47	Sequence 47, Appl
23	2322.6	94.3	3624	16	US-10-190-305A-41	Sequence 41, Appl
24	2304.4	93.6	3597	15	US-10-190-435-46	Sequence 46, Appl
25	2304.4	93.6	3597	16	US-10-190-305A-40	Sequence 40, Appl
26	2142	87.0	2466	16	US-10-241-009-31	Sequence 31, Appl
27	2142	87.0	2466	16	US-10-190-434B-31	Sequence 31, Appl
28	2142	87.0	2466	16	US-10-190-305A-33	Sequence 33, Appl
29	2122.8	86.2	2472	16	US-10-241-009-32	Sequence 32, Appl
30	2122.8	86.2	2472	16	US-10-190-434B-32	Sequence 32, Appl
31	2122.8	86.2	2472	16	US-10-190-305A-34	Sequence 34, Appl
32	2115.2	85.9	2460	16	US-10-241-009-30	Sequence 30, Appl
33	2115.2	85.9	2460	16	US-10-190-434B-30	Sequence 30, Appl
34	2115.2	85.9	2460	16	US-10-190-305A-32	Sequence 32, Appl
35	2114.2	85.8	3564	16	US-10-241-009-13	Sequence 13, Appl
36	2114.2	85.8	3564	16	US-10-241-009-14	Sequence 14, Appl
37	2114.2	85.8	3564	16	US-10-190-434B-13	Sequence 13, Appl
38	2114.2	85.8	3564	16	US-10-190-434B-14	Sequence 14, Appl
39	2113.6	85.8	4716	15	US-10-190-435-17	Sequence 17, Appl
40	2113.6	85.8	4716	16	US-10-190-305A-13	Sequence 13, Appl
41	2109.8	85.7	3999	16	US-10-241-009-9	Sequence 9, Appli
42	2109.8	85.7	3999	16	US-10-190-434B-9	Sequence 9, Appli
43	2108.2	85.6	3999	16	US-10-241-009-10	Sequence 10, Appl
44	2108.2	85.6	3999	16	US-10-241-009-11	Sequence 11, Appl
45	2108.2	85.6	3999	16	US-10-190-434B-10	Sequence 10, Appl

ALIGNMENTS

```
RESULT 1
US-09-899-575-31
; Sequence 31, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
;  APPLICANT: Zur Megede, Jan
;  APPLICANT:  Barnett, Susan W.
;  APPLICANT:  Egnelbrecht, Susan
;  APPLICANT:  van Rensburg, Estrelita Janse
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP01631.102
;  CURRENT APPLICATION NUMBER: US/09/899,575
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: 09/475,704
;  PRIOR FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 135
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 2463
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: PR975YM
US-09-899-575-31

  Query Match             100.0%;  Score 2463;  DB 10;  Length 2463;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2463;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
```

```
Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy        2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy        2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy        2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy        2461 TTC 2463
               |||
Db        2461 TTC 2463
```

```
RESULT 2
US-10-190-435-44
; Sequence 44, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-435-44

  Query Match             99.7%;  Score 2455.4;  DB 15;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2456;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy           1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy          61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy         121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy         181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy         241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy         301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy         361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy         421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
               |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy         481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy         541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy         601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy         661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy         721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy         781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy         841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy         901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy         961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy        1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy        1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy        1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy        1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy        1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy        1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy        1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy        1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 3
US-10-190-305A-38
; Sequence 38, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-305A-38

```
  Query Match             99.7%;  Score 2455.4;  DB 16;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2456;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
```

```
Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
RESULT 4
US-09-899-575-30
; Sequence 30, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
;  APPLICANT: Zur Megede, Jan
;  APPLICANT:  Barnett, Susan W.
;  APPLICANT:  Egnelbrecht, Susan
;  APPLICANT:  van Rensburg, Estrelita Janse
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP01631.102
;  CURRENT APPLICATION NUMBER: US/09/899,575
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: 09/475,704
;  PRIOR FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 135
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 2469
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: PR975(+)
US-09-899-575-30

  Query Match             99.2%;  Score 2442.2;  DB 10;  Length 2469;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
```

```
Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              |||||||||||||||||      || |  ||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2455 GGTGAATTC 2463
              |||||||||
Db       2461 GGTGAATTC 2469
```

```
RESULT 5
US-09-899-575-32
; Sequence 32, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
;  APPLICANT: Zur Megede, Jan
;  APPLICANT:  Barnett, Susan W.
;  APPLICANT:  Egnelbrecht, Susan
;  APPLICANT:  van Rensburg, Estrelita Janse
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP01631.102
;  CURRENT APPLICATION NUMBER: US/09/899,575
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: 09/475,704
;  PRIOR FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 135
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: PR975YMWM
US-09-899-575-32

  Query Match             98.9%;  Score 2436.2;  DB 10;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||    ||      ||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
```

```
Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454

Qy       2461 TTC 2463
              |||
Db       2455 TTC 2457
```

```
RESULT 6
US-10-190-435-45
; Sequence 45, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-435-45

  Query Match             98.7%;  Score 2430.2;  DB 15;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2448;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660
```

```
Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              |||||||||||      || |  ||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
RESULT 7
US-10-190-305A-39
; Sequence 39, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-305A-39

  Query Match             98.7%;  Score 2430.2;  DB 16;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 0;
```

```
Matches 2448;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              |||||||||||      || |  |||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
```

```
Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 8
US-10-190-435-43
; Sequence 43, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;   LENGTH: 2445
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMWM_C
US-10-190-435-43

  Query Match             98.4%;  Score 2422.6;  DB 15;  Length 2445;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2441;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              ||||||||||||||||||||||||   ||      |||||||||||||||||||||||||
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACC 1374

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1434
```

```
Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1494

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1554

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1614

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1674

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1794

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1854

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1914

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1974

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2034

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2094

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2154

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2214

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2274

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2334

Qy       2347 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2394

Qy       2407 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 9
US-10-190-305A-37
; Sequence 37, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: p2Pol.opt.YMWM_C
US-10-190-305A-37

```
  Query Match             98.4%;  Score 2422.6;  DB 16;  Length 2445;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2441;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy         787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy         847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy         907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy         967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy        1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy        1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy        1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy        1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy        1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy        1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
               ||||||||||||||||||||||||||||    ||      ||||||||||||||||||||||
Db        1321 CACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACC 1374

Qy        1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1375 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1434

Qy        1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1435 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1494

Qy        1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1495 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1554

Qy        1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1555 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1614

Qy        1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1615 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1674

Qy        1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1675 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734

Qy        1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1735 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1794

Qy        1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1795 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1854

Qy        1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1855 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1914

Qy        1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1915 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1974

Qy        1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1975 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2034

Qy        2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2035 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2094

Qy        2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2095 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2154

Qy        2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2155 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2214

Qy        2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2215 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2274

Qy        2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2275 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2334

Qy        2347 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2406
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2335 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2394

Qy        2407 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2395 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445

RESULT 10
US-10-190-435-9
; Sequence 9, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmut_C
US-10-190-435-9

  Query Match             98.1%;  Score 2415.6;  DB 15;  Length 3930;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2434;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
```

```
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy        74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy       134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy       194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy       254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy       314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy       374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy       434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy       494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy       554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy       614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy       674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy       734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy      1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy      1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
             |||||||||||||||||   ||      ||||||||||||||||||||||||||||||||
Db      2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy      1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy      1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy      1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy      1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy      1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy      1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy      1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy      1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy      1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy      1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy      1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy      2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy      2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy      2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700
```

```
Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              ||||||||||||||||||||||||||||||||||||||||||||
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 11
US-10-190-435-10
; Sequence 10, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt_C
US-10-190-435-10

  Query Match             98.0%;  Score 2414;  DB 15;  Length 3930;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2433;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
              |||||||||||||||||   ||      ||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980
```

```
Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              ||||||||||||||||||||||||||||||||||||||||||||
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 12
US-10-190-435-11
; Sequence 11, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna_C
US-10-190-435-11

```
  Query Match             98.0%;  Score 2414;  DB 15;  Length 3930;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2433;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
```

```
Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
              |||||||||||||||||   ||      ||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              ||||||||||||||||||||||||||||||||||||||||||||
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 13
US-10-190-435-58
; Sequence 58, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
;   LENGTH: 5184
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: TatRevNefgagCpolIna C
US-10-190-435-58

  Query Match             98.0%;  Score 2414;  DB 15;  Length 5184;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2433;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800
```

```
Qy        74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy       134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy       194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy       254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy       314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy       374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy       434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
             ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy       494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy       554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy       614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy       674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy       734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy      1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000

Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060

Qy      1334 AGGAGCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
             ||||||||||||||||   ||      |||||||||||||||||||||||||||||||||
Db      4061 AGGAGCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 4114

Qy      1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4115 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 4174

Qy      1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4175 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 4234

Qy      1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4235 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 4294

Qy      1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4295 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 4354

Qy      1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4355 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 4414

Qy      1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4415 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 4474

Qy      1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4475 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 4534

Qy      1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4535 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 4594

Qy      1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4595 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 4654

Qy      1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4655 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 4714

Qy      1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4715 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 4774

Qy      2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4775 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 4834

Qy      2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4835 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 4894

Qy      2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4895 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 4954

Qy      2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4955 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 5014

Qy     2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5015 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 5074

Qy     2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5075 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 5134

Qy     2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
            ||||||||||||||||||||||||||||||||||||||||||||
Db     5135 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 14
US-10-190-305A-82
; Sequence 82, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
;   LENGTH: 5184
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: TatRevNefgagCpolIna C
US-10-190-305A-82

```
  Query Match             98.0%;  Score 2414;  DB 16;  Length 5184;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2433;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy       14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy       74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy      134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy      194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy      254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy      314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy      374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy      434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
            ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy      494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy      554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy      614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy      674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy      734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy      794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy      854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy      914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy      974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy     1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy     1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy     1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy     1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000

Qy     1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060

Qy     1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
            |||||||||||||||||   ||      ||||||||||||||||||||||||||||||||
Db     4061 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 4114

Qy     1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4115 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 4174

Qy     1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4175 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 4234

Qy     1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4235 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 4294

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4295 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 4354

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4355 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 4414

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4415 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 4474

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4475 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 4534

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4535 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 4594

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4595 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 4654

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4655 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 4714

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4715 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 4774

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4775 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 4834

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4835 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 4894

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4895 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 4954

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4955 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 5014

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5015 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 5074

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5075 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 5134

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              ||||||||||||||||||||||||||||||||||||||||||||
Db       5135 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 15
US-10-190-435-13
; Sequence 13, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 3531
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_C
US-10-190-435-13

```
  Query Match             96.8%;  Score 2383.6;  DB 15;  Length 3531;
  Best Local Similarity   98.8%;  Pred. No. 0;
  Matches 2414;  Conservative    0;  Mismatches   24;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||| ||||||||||||| || |   |    ||   ||   | ||  ||| |       |
Db       1088 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1147

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1207

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1267

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1327

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1387

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1447

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1507

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1508 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1567

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1568 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1627

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1628 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1687

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1688 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1747

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1748 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1807

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1808 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1867

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        1868 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1927

Qy         854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1928 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987

Qy         914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2047

Qy         974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2048 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2107

Qy        1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2108 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2167

Qy        1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2168 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2227

Qy        1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2228 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2287

Qy        1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2288 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2347

Qy        1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2348 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2407

Qy        1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
               ||||||||||||||||    ||       |||||||||||||||||||||||||||||||
Db        2408 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2461

Qy        1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2462 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2521

Qy        1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2522 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2581

Qy        1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2582 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 2641

Qy        1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2642 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 2701

Qy        1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2702 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 2761

Qy        1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2762 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 2821

Qy        1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2822 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 2881

Qy        1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2882 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 2941

Qy        1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2942 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3001

Qy        1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3002 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3061

Qy        1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3062 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3121

Qy        2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3122 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3181

Qy        2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3182 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3241

Qy        2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3242 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3301

Qy        2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3302 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3361

Qy        2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3362 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3421

Qy        2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3422 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3481

Qy        2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
               ||||||||||||||||||||||||||||||||||||||||||||
Db        3482 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3525

Search completed: June  2, 2005, 04:23:40
Job time : 1419.33 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:          June  1, 2005, 03:50:55 ; Search time 10513 Seconds
                                           (without alignments)
                                           11324.506 Million cell updates/sec

Title:           US-09-610-313B-32
Perfect score:   2457
Sequence:        1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2457

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters:      9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_htg:*
                 3:  gb_in:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pl:*
                 9:  gb_pr:*
                 10:  gb_ro:*
                 11:  gb_sts:*
                 12:  gb_sy:*
                 13:  gb_un:*
                 14:  gb_vi:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2457	100.0	2457	6	AX455916	AX455916 Sequence
2	2436.2	99.2	2463	6	AX455915	AX455915 Sequence
3	2415.4	98.3	2469	6	AX455914	AX455914 Sequence
4	2040	83.0	2300	6	BD263705	BD263705 Improved
5	2040	83.0	2300	6	CQ870575	CQ870575 Sequence
6	2040	83.0	2300	6	AR373388	AR373388 Sequence
7	2019.2	82.2	2306	6	BD263704	BD263704 Improved
8	2019.2	82.2	2306	6	CQ870574	CQ870574 Sequence
9	2019.2	82.2	2306	6	AR373387	AR373387 Sequence
10	2000.8	81.4	9166	6	AX427930	AX427930 Sequence
11	1998.4	81.3	2312	6	BD263706	BD263706 Improved
12	1998.4	81.3	2312	6	CQ870576	CQ870576 Sequence
13	1998.4	81.3	2312	6	AR373389	AR373389 Sequence
14	1992.4	81.1	9788	6	AX427936	AX427936 Sequence
15	1978.4	80.5	9169	6	AX427931	AX427931 Sequence
16	1975	80.4	9194	6	AX427926	AX427926 Sequence
17	1973.4	80.3	9194	6	AX427925	AX427925 Sequence
18	1973.4	80.3	12411	6	AX427927	AX427927 Sequence
19	1970	80.2	9785	6	AX427938	AX427938 Sequence
20	1966.6	80.0	9167	6	AX427933	AX427933 Sequence
21	1966.6	80.0	9170	6	AX427928	AX427928 Sequence
22	1966.6	80.0	9782	6	AX427935	AX427935 Sequence
23	1966.6	80.0	9783	6	AX427934	AX427934 Sequence
24	1966.6	80.0	9792	6	AX427932	AX427932 Sequence
25	1965	80.0	9189	6	AX427921	AX427921 Sequence
26	1962	79.9	4353	12	AF287353	AF287353 Synthetic
27	1945.4	79.2	3015	6	AX455946	AX455946 Sequence
28	1928.8	78.5	3009	6	AX455987	AX455987 Sequence
29	1915.8	78.0	4319	6	BD263639	BD263639 Improved
30	1915.8	78.0	4319	6	CQ870498	CQ870498 Sequence
31	1915.8	78.0	4319	6	AR373322	AR373322 Sequence
32	1905.2	77.5	4352	12	AF287352	AF287352 Synthetic
33	1878.6	76.5	1965	6	AX455952	AX455952 Sequence
34	1872.8	76.2	2299	6	BD263703	BD263703 Improved
35	1872.8	76.2	2299	6	CQ870573	CQ870573 Sequence
36	1872.8	76.2	2299	6	AR373386	AR373386 Sequence
37	1859.8	75.7	2577	6	AX457088	AX457088 Sequence
38	1852.2	75.4	4343	6	AX306428	AX306428 Sequence
39	1852	75.4	2305	6	BD263702	BD263702 Improved
40	1852	75.4	2305	6	CQ870572	CQ870572 Sequence
41	1852	75.4	2305	6	AR373385	AR373385 Sequence
42	1848.2	75.2	4341	6	AX306429	AX306429 Sequence
43	1838.6	74.8	1977	6	AX455954	AX455954 Sequence
44	1825.6	74.3	7897	6	AX427923	AX427923 Sequence
45	1772	72.1	3024	12	AF287355	AF287355 Synthetic

ALIGNMENTS

```
RESULT 1
AX455916
LOCUS       AX455916                2457 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 32 from Patent WO0204493.
ACCESSION   AX455916
VERSION     AX455916.1  GI:21714901
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 32 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1..2457
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YMWM"
ORIGIN

  Query Match            100.0%;  Score 2457;  DB 6;  Length 2457;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-247;
  Matches 2457;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340

Qy       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400
```

```
Qy       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457

RESULT 2
AX455915
LOCUS       AX455915                2463 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 31 from Patent WO0204493.
ACCESSION   AX455915
VERSION     AX455915.1  GI:21714900
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 31 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1..2463
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YM"
ORIGIN

  Query Match             99.2%;  Score 2436.2;  DB 6;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 2.1e-245;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              ||||||||||||||||||||||||||||    ||      |||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2455 TTC 2457
              |||
Db       2461 TTC 2463

RESULT 3
AX455914
LOCUS       AX455914                2469 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 30 from Patent WO0204493.
ACCESSION   AX455914
VERSION     AX455914.1  GI:21714899
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 30 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1..2469
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975(+)"
ORIGIN

  Query Match            98.3%;  Score 2415.4;  DB 6;  Length 2469;
  Best Local Similarity  99.3%;  Pred. No. 3.2e-243;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368
              ||||||||||||||||||||||||||||||||||||   ||      |||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2449 GGTGAATTC 2457
              |||||||||
Db       2461 GGTGAATTC 2469
```

```
RESULT 4
BD263705
LOCUS       BD263705                2300 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263705
VERSION     BD263705.1  GI:33073473
KEYWORDS    JP 2002533124-A/72.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2300)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 72 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/72
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YMWM FH
            Key             Location/Qualifiers
            FT   source          1..2300
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1..2300
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN
```

```
Query Match             83.0%;  Score 2040;  DB 6;  Length 2300;
Best Local Similarity   93.6%;  Pred. No. 4.6e-204;
Matches 2153;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1200

Qy       1358 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGA 1417
Db       1201 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA 1260

Qy       1418 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477
Db       1261 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG 1320

Qy       1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG 1537
Db       1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGA 1380

Qy       1538 TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGC 1597
Db       1381 TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC 1440

Qy       1598 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1657
Db       1441 CCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1500

Qy       1658 GCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1717
Db       1501 GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1560

Qy       1718 AGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1777
Db       1561 AGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1620

Qy       1778 AGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCA 1837
Db       1621 AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA 1680

Qy       1838 TCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCG 1897
Db       1681 TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG 1740

Qy       1898 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1957
Db       1741 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1800

Qy       1958 CCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCG 2017
Db       1801 CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG 1860

Qy       2018 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCA 2077
Db       1861 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA 1920

Qy       2078 CCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG 2137
Db       1921 CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG 1980

Qy       2138 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2197
Db       1981 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2040

Qy       2198 GCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2257
Db       2041 GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2100

Qy       2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
```

```
Db       2101 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2160

Qy       2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377
Db       2161 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2220

Qy       2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437
Db       2221 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2280

Qy       2438 GGGCTAGCACCGGTGAATTC 2457
Db       2281 GGGCTAGCACCGGTGAATTC 2300
```

```
RESULT 5
CQ870575
LOCUS       CQ870575                2300 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 83 from Patent EP1433851.
ACCESSION   CQ870575
VERSION     CQ870575.1  GI:52000091
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 83 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..2300
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt.YMWM"
ORIGIN

  Query Match             83.0%;  Score 2040;  DB 6;  Length 2300;
  Best Local Similarity   93.6%;  Pred. No. 4.6e-204;
  Matches 2153;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1200

Qy       1358 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGA 1417
Db       1201 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA 1260

Qy       1418 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477
Db       1261 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG 1320

Qy       1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG 1537
Db       1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGA 1380

Qy       1538 TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGC 1597
Db       1381 TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC 1440

Qy       1598 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1657
Db       1441 CCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1500
```

```
Qy     1658 GCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1717
Db     1501 GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1560

Qy     1718 AGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1777
Db     1561 AGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1620

Qy     1778 AGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCA 1837
Db     1621 AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA 1680

Qy     1838 TCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCG 1897
Db     1681 TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG 1740

Qy     1898 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1957
Db     1741 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1800

Qy     1958 CCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCG 2017
Db     1801 CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG 1860

Qy     2018 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCA 2077
Db     1861 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA 1920

Qy     2078 CCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG 2137
Db     1921 CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG 1980

Qy     2138 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2197
Db     1981 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2040

Qy     2198 GCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2257
Db     2041 GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2100

Qy     2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
Db     2101 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2160

Qy     2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377
Db     2161 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2220

Qy     2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437
Db     2221 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2280

Qy     2438 GGGCTAGCACCGGTGAATTC 2457
Db     2281 GGGCTAGCACCGGTGAATTC 2300
```

```
RESULT 6
AR373388
LOCUS       AR373388              2300 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 83 from patent US 6602705.
ACCESSION   AR373388
VERSION     AR373388.1  GI:40075491
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2300)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 83 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1..2300
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match            83.0%;  Score 2040;  DB 6;  Length 2300;
  Best Local Similarity  93.6%;  Pred. No. 4.6e-204;
  Matches 2153;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy      170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db        1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy      230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db       61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy      290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db      121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy      344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db      181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy      398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db      241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy      458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db      301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy      518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db      361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy      578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db      421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy      638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db      481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy      698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db      541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy      758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db      601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy      818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db      661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy      878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db      721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy      938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db      781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy      998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db      841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy     1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db      901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960
```

```
Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1200

Qy       1358 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGA 1417
Db       1201 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA 1260

Qy       1418 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477
Db       1261 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG 1320

Qy       1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG 1537
Db       1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGA 1380

Qy       1538 TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGC 1597
Db       1381 TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC 1440

Qy       1598 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1657
Db       1441 CCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1500

Qy       1658 GCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1717
Db       1501 GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1560

Qy       1718 AGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1777
Db       1561 AGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1620

Qy       1778 AGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCA 1837
Db       1621 AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA 1680

Qy       1838 TCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCG 1897
Db       1681 TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG 1740

Qy       1898 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1957
Db       1741 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1800

Qy       1958 CCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCG 2017
Db       1801 CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG 1860

Qy       2018 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCA 2077
Db       1861 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA 1920

Qy       2078 CCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG 2137
Db       1921 CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG 1980

Qy       2138 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2197
Db       1981 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2040

Qy       2198 GCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2257
Db       2041 GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2100

Qy       2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
Db       2101 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2160

Qy       2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377
Db       2161 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2220

Qy       2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437
Db       2221 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2280

Qy       2438 GGGCTAGCACCGGTGAATTC 2457
Db       2281 GGGCTAGCACCGGTGAATTC 2300

RESULT 7
BD263704
LOCUS       BD263704                2306 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263704
VERSION     BD263704.1  GI:33073472
KEYWORDS    JP 2002533124-A/71.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 71 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/71
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH   Key
                       Location/Qualifiers
            FT   source          1..2306
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1..2306
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             82.2%;  Score 2019.2;  DB 6;  Length 2306;
  Best Local Similarity   93.2%;  Pred. No. 6.8e-202;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
```

```
Db    121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy    344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db    181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy    398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db    241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy    458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db    301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy    518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db    361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy    578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db    421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy    638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db    481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy    698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db    541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy    758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db    601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy    818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db    661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy    878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db    721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy    938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db    781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy    998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db    841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy   1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db    901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy   1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db    961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy   1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db   1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy   1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db   1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy   1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1355
Db   1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy   1356 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1411
Db   1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy   1412 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1471
Db   1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy   1472 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1531
Db   1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy   1532 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1591
Db   1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy   1592 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1651
Db   1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy   1652 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1711
Db   1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy   1712 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1771
Db   1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy   1772 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1831
Db   1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy   1832 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1891
Db   1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy   1892 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1951
Db   1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy   1952 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2011
Db   1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy   2012 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2071
Db   1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy   2072 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2131
Db   1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy   2132 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2191
Db   1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy   2192 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2251
Db   2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy   2252 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2311
Db   2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy   2312 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2371
Db   2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy   2372 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2431
Db   2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy   2432 TTCCCGGGGCTAGCACCGGTGAATTC 2457
Db   2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

```
RESULT 8
CQ870574
LOCUS       CQ870574                 2306 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 82 from Patent EP1433851.
ACCESSION   CQ870574
VERSION     CQ870574.1  GI:52000090
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 82 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..2306
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt.YM"
ORIGIN

  Query Match             82.2%;  Score 2019.2;  DB 6;  Length 2306;
  Best Local Similarity   93.2%;  Pred. No. 6.8e-202;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1355
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy       1356 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1411
Db       1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy       1412 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1471
Db       1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy       1472 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1531
Db       1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy       1532 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1591
Db       1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy       1592 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1651
Db       1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy       1652 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1711
Db       1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy       1712 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1771
Db       1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy       1772 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1831
Db       1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy       1832 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1891
```

```
Db       1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy       1892 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1951

Db       1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy       1952 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2011

Db       1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy       2012 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2071

Db       1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy       2072 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2131

Db       1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy       2132 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2191

Db       1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy       2192 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2251

Db       2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy       2252 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2311

Db       2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy       2312 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2371

Db       2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy       2372 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2431

Db       2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy       2432 TTCCCGGGGCTAGCACCGGTGAATTC 2457

Db       2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

```
RESULT 9
AR373387
LOCUS       AR373387                2306 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 82 from patent US 6602705.
ACCESSION   AR373387
VERSION     AR373387.1  GI:40075490
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 82 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1..2306
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             82.2%;  Score 2019.2;  DB 6;  Length 2306;
  Best Local Similarity   93.2%;  Pred. No. 6.8e-202;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757

Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297

Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1355
```

```
Db    1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy    1356 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1411
Db    1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy    1412 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1471
Db    1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy    1472 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1531
Db    1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy    1532 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1591
Db    1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy    1592 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1651
Db    1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy    1652 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1711
Db    1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy    1712 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1771
Db    1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy    1772 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1831
Db    1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy    1832 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1891
Db    1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy    1892 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1951
Db    1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy    1952 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2011
Db    1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy    2012 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2071
Db    1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy    2072 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2131
Db    1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy    2132 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2191
Db    1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy    2192 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2251
Db    2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy    2252 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2311
Db    2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy    2312 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2371
Db    2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy    2372 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2431
Db    2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy    2432 TTCCCGGGGCTAGCACCGGTGAATTC 2457
Db    2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

```
RESULT 10
AX427930
LOCUS       AX427930                9166 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 168 from Patent WO0232943.
ACCESSION   AX427930
VERSION     AX427930.1  GI:21538017
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 168 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9166
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             81.4%;  Score 2000.8;  DB 6;  Length 9166;
  Best Local Similarity   91.3%;  Pred. No. 3.9e-200;
  Matches 2185;  Conservative    0;  Mismatches  182;  Indels   25;  Gaps    5;

Qy      14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db    2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy      74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db    3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy     134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db    3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy     194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db    3141 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3199

Qy     254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db    3200 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3259

Qy     314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db    3260 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3319

Qy     365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db    3320 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3379

Qy     422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db    3380 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3439

Qy     482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db    3440 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3499

Qy     542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db    3500 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3559
```

```
Qy    602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db   3560 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3619

Qy    662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db   3620 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3679

Qy    722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db   3680 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3739

Qy    782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db   3740 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3799

Qy    842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db   3800 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3859

Qy    902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db   3860 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3919

Qy    962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db   3920 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3979

Qy   1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db   3980 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4039

Qy   1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db   4040 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4099

Qy   1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db   4100 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4159

Qy   1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db   4160 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4219

Qy   1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db   4220 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4279

Qy   1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCG 1369
Db   4280 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4339

Qy   1370 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1429
Db   4340 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4399

Qy   1430 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1489
Db   4400 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4459

Qy   1490 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1549
Db   4460 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4519

Qy   1550 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1609
Db   4520 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4579

Qy   1610 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1669
Db   4580 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4639

Qy   1670 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1729
Db   4640 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4699

Qy   1730 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1789
Db   4700 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4759

Qy   1790 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1849
Db   4760 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4819

Qy   1850 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1909
Db   4820 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4879

Qy   1910 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1969
Db   4880 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4939

Qy   1970 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029
Db   4940 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4999

Qy   2030 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2089
Db   5000 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5059

Qy   2090 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2149
Db   5060 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5119

Qy   2150 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2209
Db   5120 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5179

Qy   2210 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2269
Db   5180 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5239

Qy   2270 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2329
Db   5240 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5299

Qy   2330 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
Db   5300 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5351
```

```
RESULT 11
BD263706
LOCUS       BD263706                2312 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263706
VERSION     BD263706.1  GI:33073474
KEYWORDS    JP 2002533124-A/73.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 73 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/73
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
```

```
          KARIN HARTOG,
          PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
          C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
          C12N5/10,
          PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
          Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH    Key
                   Location/Qualifiers
          FT   source          1. .2312
          FT                   /organism='Artificial Sequence'.
FEATURES            Location/Qualifiers
     source         1. .2312
                    /organism="synthetic construct"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:32630"
ORIGIN

  Query Match          81.3%;  Score 1998.4;  DB 6;  Length 2312;
  Best Local Similarity   92.9%;  Pred. No. 1e-199;
  Matches 2147;  Conservative    0;  Mismatches  141;  Indels   24;  Gaps    4;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGC 1351
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1405
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1406 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1465
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1466 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1525
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1526 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1585
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1586 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1645
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1646 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1705
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1706 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1765
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1766 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1825
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1826 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1885
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1886 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1945
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1946 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2005
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2006 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2065
```

```
Db     1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy     2066 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2125

Db     1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy     2126 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2185

Db     1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy     2186 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2245

Db     2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy     2246 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2305

Db     2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy     2306 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2365

Db     2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy     2366 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2425

Db     2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy     2426 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457

Db     2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 12
CQ870576
LOCUS       CQ870576                2312 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 84 from Patent EP1433851.
ACCESSION   CQ870576
VERSION     CQ870576.1  GI:52000092
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 84 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..2312
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt(+)"
ORIGIN

  Query Match            81.3%;  Score 1998.4;  DB 6;  Length 2312;
  Best Local Similarity  92.9%;  Pred. No. 1e-199;
  Matches 2147;  Conservative    0;  Mismatches  141;  Indels   24;  Gaps    4;

Qy      170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db        1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy      230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db       61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy      290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db      121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy      344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db      181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy      398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db      241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy      458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db      301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy      518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db      361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy      578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db      421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy      638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db      481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy      698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757

Db      541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy      758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db      601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy      818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db      661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy      878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db      721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy      938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db      781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy      998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db      841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy     1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db      901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy     1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db      961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy     1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231

Db     1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy     1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291

Db     1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy     1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGC 1351

Db     1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy     1352 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1405

Db     1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy     1406 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1465
```

```
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1466 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1525
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1526 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1585
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1586 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1645
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1646 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1705
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1706 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1765
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1766 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1825
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1826 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1885
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1886 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1945
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1946 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2005
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2006 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2065
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2066 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2125
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2126 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2185
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2186 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2245
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2246 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2305
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2306 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2365
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2366 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2425
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2426 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

RESULT 13

```
AR373389
LOCUS       AR373389                2312 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 84 from patent US 6602705.
ACCESSION   AR373389
VERSION     AR373389.1  GI:40075492
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 84 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1..2312
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match          81.3%;  Score 1998.4;  DB 6;  Length 2312;
  Best Local Similarity   92.9%;  Pred. No. 1e-199;
  Matches 2147;  Conservative    0;  Mismatches  141;  Indels   24;  Gaps    4;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
              ||||||||||    |  ||||||| |||||||||||||         || |  |||||||
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
              |||||||||||||||||||||||||||||||| ||||||||||||||||||| ||||||
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGC 1351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1405
                ||      ||||||||||||||||||||||||||||||||||||  ||||||||||
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1406 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1465
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1466 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1525
              ||||||| ||||||||||||||   |||||||||||||||||||||||| |||||||||
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1526 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1585
              ||||||  |  |||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1586 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1645
              |||||   |||||||||||| |||||||||||||||||||||||||||||||||||||||
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1646 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1705
              |||||||||||||| | |||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1706 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1765
              |||||||||||||||||||    ||||||  |||||||||||||||||||||||||||||
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1766 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1825
              |||||||||||||||| |  ||   ||||||||||||||||||||||||||| ||||||||
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1826 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1885
              ||  |||||||||||||||||||||||||||| |||||||| ||  ||||||||||||||
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1886 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1945
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1946 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2005
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2006 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2065
              ||||||| | ||||||||||||||||||||||||||||||||||||||| | ||||||
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2066 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2125
              |  |||| |||||||||||||||||||||||||||||||||||| |||||||||||||||
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2126 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2185
              |||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2186 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2245
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2246 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2305
              |||||||||||||  |||||||||||||||||||||||||||||||||||||||| | |
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2306 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2365
              |||||||||||||   ||||||||||||||||||||||||| ||||||||||||||||||
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2366 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2426 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
              ||||||||||||||||||||||||||||||||
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312


RESULT 14
AX427936
LOCUS       AX427936                9788 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 174 from Patent WO0232943.
ACCESSION   AX427936
VERSION     AX427936.1  GI:21538023
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 174 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9788
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             81.1%;  Score 1992.4;  DB 6;  Length 9788;
  Best Local Similarity   91.2%;  Pred. No. 2.9e-199;
  Matches 2181;  Conservative    0;  Mismatches  181;  Indels   30;  Gaps    5;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              ||||||||||||||| |||  | ||  | ||||||||  ||||||||||||||||||| ||
Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              |||||||||||||| ||||||||||||||||||||||||||||||||||||  | ||||||
Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
```

```
Db   3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy    194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db   3141 AGGACTGCACCGAGCGACAGGCTAA------TAGGGAAGATCTGGCCTTCCCACAAGGGA 3194

Qy    254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db   3195 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3254

Qy    314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db   3255 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3314

Qy    365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db   3315 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3374

Qy    422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db   3375 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3434

Qy    482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db   3435 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3494

Qy    542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db   3495 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3554

Qy    602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db   3555 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3614

Qy    662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db   3615 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3674

Qy    722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db   3675 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3734

Qy    782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db   3735 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3794

Qy    842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db   3795 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3854

Qy    902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db   3855 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3914

Qy    962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db   3915 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3974

Qy   1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db   3975 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4034

Qy   1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db   4035 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4094

Qy   1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db   4095 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4154

Qy   1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db   4155 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4214

Qy   1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db   4215 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4274

Qy   1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCG 1369
Db   4275 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4334

Qy   1370 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1429
Db   4335 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4394

Qy   1430 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1489
Db   4395 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4454

Qy   1490 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1549
Db   4455 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4514

Qy   1550 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1609
Db   4515 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4574

Qy   1610 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1669
Db   4575 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4634

Qy   1670 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1729
Db   4635 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4694

Qy   1730 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1789
Db   4695 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4754

Qy   1790 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1849
Db   4755 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4814

Qy   1850 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1909
Db   4815 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4874

Qy   1910 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1969
Db   4875 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4934

Qy   1970 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029
Db   4935 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4994

Qy   2030 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2089
Db   4995 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5054

Qy   2090 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2149
Db   5055 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5114

Qy   2150 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2209
Db   5115 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5174

Qy   2210 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2269
Db   5175 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5234

Qy   2270 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2329
Db   5235 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5294
```

```
Qy       2330 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
Db       5295 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5346

RESULT 15
AX427931
LOCUS       AX427931                9169 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 169 from Patent WO0232943.
ACCESSION   AX427931
VERSION     AX427931.1  GI:21538018
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 169 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9169
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             80.5%;  Score 1978.4;  DB 6;  Length 9169;
  Best Local Similarity   91.1%;  Pred. No. 8.5e-198;
  Matches 2178;  Conservative    0;  Mismatches  186;  Indels   28;  Gaps    6;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2967 TGGCCGAGGCCATGAGCCAGGTGCAGCACACCAACATCATGATGCAGCGCGGCAACTTCA 3026

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       3027 AGGGCCAGAAGCGCAT---CAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3083

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       3084 ACTGCCGCGCCCTGCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 3143

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       3144 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3202

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       3203 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3262

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db       3263 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3322

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db       3323 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3382

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db       3383 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3442

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db       3443 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3502

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db       3503 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3562

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db       3563 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3622

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db       3623 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3682

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db       3683 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3742

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db       3743 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3802

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db       3803 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3862

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db       3863 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3922

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db       3923 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3982

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db       3983 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4042

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db       4043 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4102

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db       4103 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4162

Qy       1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db       4163 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4222

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db       4223 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4282

Qy       1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCG 1369
Db       4283 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4342

Qy       1370 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1429
Db       4343 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4402

Qy       1430 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1489
Db       4403 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4462

Qy       1490 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1549
Db       4463 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4522

Qy       1550 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1609
Db       4523 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4582

Qy       1610 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1669
Db       4583 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4642

Qy       1670 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1729
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||     |
Db        4643 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4702

Qy        1730 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1789
               ||     |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4703 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4762

Qy        1790 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1849
                 ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||
Db        4763 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4822

Qy        1850 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1909
               |||||||| ||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        4823 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4882

Qy        1910 TGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1969
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4883 TGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4942

Qy        1970 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029
               |||||||||||||||||||||||||||||||||||||||||||| | |||||||||||||
Db        4943 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 5002

Qy        2030 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2089
               ||||||||||||| ||||| |||||| | |||  ||||||||| ||||||||||||||||
Db        5003 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5062

Qy        2090 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2149
               ||||||||||||||||||| |||||||||||||||||||||   ||||||||||||||||
Db        5063 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5122

Qy        2150 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2209
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5123 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5182

Qy        2210 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2269
               |||||| |||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db        5183 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5242

Qy        2270 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2329
               |||||||||||||||||||||||||||||||| | |||   |||||||||   |||||||
Db        5243 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5302

Qy        2330 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
               ||||||||||||||||||||||||||||    | |   | |    ||  || |
Db        5303 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5354
```

Search completed: June 1, 2005, 20:20:49
Job time : 10525 secs

OM nucleic - nucleic search, using sw model

Run on: May 31, 2005, 23:20:00 ; Search time 1264.25 Seconds
(without alignments)
11504.726 Million cell updates/sec

Title: US-09-610-313B-32
Perfect score: 2457
Sequence: 1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2457	100.0	2457	6	ABL39961	Abl39961 Synthetic
2	2457	100.0	2457	12	ADM73766	Adm73766 HIV-1 pol
3	2443.4	99.4	2445	8	ACA03546	Aca03546 Synthetic
4	2443.4	99.4	2445	10	ADC13264	Adc13264 DNA of HI
5	2436.4	99.2	3930	10	ADC13230	Adc13230 DNA of HI
6	2436.2	99.2	2463	6	ABL39960	Abl39960 Synthetic
7	2436.2	99.2	2463	12	ADM73765	Adm73765 HIV-1 pol
8	2434.8	99.1	3930	10	ADC13231	Adc13231 DNA of HI
9	2434.8	99.1	3930	10	ADC13232	Adc13232 DNA of HI
10	2434.8	99.1	5184	8	ACA03591	Aca03591 Synthetic
11	2434.8	99.1	5184	10	ADC13279	Adc13279 DNA of HI
12	2428.6	98.8	2457	8	ACA03547	Aca03547 Synthetic
13	2428.6	98.8	2457	10	ADC13265	Adc13265 DNA of HI
14	2415.4	98.3	2469	6	ABL39959	Abl39959 Synthetic
15	2415.4	98.3	2469	12	ADM73764	Adm73764 HIV-1 pol
16	2404.4	97.9	3531	10	ADC13234	Adc13234 DNA of HI
17	2403.4	97.8	2457	8	ACA03548	Aca03548 Synthetic
18	2403.4	97.8	2457	10	ADC13266	Adc13266 DNA of HI
19	2402.8	97.8	3537	10	ADC13236	Adc13236 DNA of HI
20	2401.8	97.8	5145	8	ACA03521	Aca03521 Synthetic
21	2401.8	97.8	5145	10	ADC13233	Adc13233 DNA of HI
22	2391.8	97.3	3538	10	ADC13235	Adc13235 DNA of HI
23	2343.4	95.4	3607	8	ACA03551	Aca03551 Synthetic
24	2343.4	95.4	3607	10	ADC13269	Adc13269 DNA of HI
25	2325.2	94.6	3597	8	ACA03549	Aca03549 Synthetic
26	2325.2	94.6	3597	10	ADC13267	Adc13267 DNA of HI
27	2295.8	93.4	3624	8	ACA03550	Aca03550 Synthetic
28	2295.8	93.4	3624	10	ADC13268	Adc13268 DNA of HI
29	2136	86.9	2460	8	ACA03541	Aca03541 Synthetic
30	2136	86.9	2460	8	ACC78505	Acc78505 HIV p2Pol
31	2135	86.9	3564	8	ACC78488	Acc78488 HIV GagPo
32	2135	86.9	3564	8	ACC78489	Acc78489 HIV GagPo
33	2134.4	86.9	4716	8	ACA03522	Aca03522 Synthetic
34	2134.4	86.9	4716	10	ADC13238	Adc13238 DNA of HI
35	2130.6	86.7	3999	8	ACC78484	Acc78484 HIV GagCo
36	2129	86.7	3999	8	ACC78485	Acc78485 HIV GagCo
37	2129	86.7	3999	8	ACC78486	Acc78486 HIV GagCo
38	2129	86.7	5283	8	ACA03584	Aca03584 Synthetic
39	2129	86.7	5283	8	ACC78529	Acc78529 HIV TatRe
40	2128.8	86.6	4713	8	ACA03592	Aca03592 Synthetic
41	2128.8	86.6	4713	10	ADC13280	Adc13280 DNA of HI
42	2128.4	86.6	3462	10	ADC13237	Adc13237 DNA of HI
43	2115.2	86.1	2466	8	ACA03542	Aca03542 Synthetic
44	2115.2	86.1	2466	8	ACC78506	Acc78506 HIV p2Pol
45	2103	85.6	3735	8	ACA03545	Aca03545 Synthetic

ALIGNMENTS

RESULT 1
ABL39961
ID ABL39961 standard; DNA; 2457 BP.
XX
AC ABL39961;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975YMWM SEQ ID NO:32.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 10; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.

```
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 2457;  DB 6;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-296;
  Matches 2457;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340

Qy       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400

Qy       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457


RESULT 2
ADM73766
ID   ADM73766 standard; DNA; 2457 BP.
XX
AC   ADM73766;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #9.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 32; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 2457;  DB 12;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-296;
  Matches 2457;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
```

```
Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340

Qy       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400

Qy       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
```

```
RESULT 3
ACA03546
ID   ACA03546 standard; DNA; 2445 BP.
XX
AC   ACA03546;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #29.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
```

```
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 34; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match             99.4%;  Score 2443.4;  DB 8;  Length 2445;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-294;
  Matches 2444;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1380

Qy       1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1440

Qy       1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500

Qy       1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1560

Qy       1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1620

Qy       1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1680

Qy       1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1740

Qy       1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1800

Qy       1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1801 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1860

Qy       1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1920

Qy       1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1980

Qy       1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2040

Qy       2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2100

Qy       2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2160

Qy       2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2220

Qy       2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2280

Qy       2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2340

Qy       2347 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2400

Qy       2407 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              |||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

```
RESULT 4
ADC13264
ID   ADC13264 standard; DNA; 2445 BP.
XX
AC   ADC13264;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YMWM_C SEQ ID NO 43.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 40; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match             99.4%;  Score 2443.4;  DB 10;  Length 2445;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-294;
  Matches 2444;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy    727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy    787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy    847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy    907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy    967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy   1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy   1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy   1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy   1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy   1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy   1327 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1380

Qy   1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1440

Qy   1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500

Qy   1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1560

Qy   1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1620

Qy   1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1680

Qy   1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1740

Qy   1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1800

Qy   1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1860

Qy   1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1920

Qy   1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1980

Qy   1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2040

Qy   2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2100

Qy   2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2160

Qy   2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2220

Qy   2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2280

Qy   2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2340

Qy   2347 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2400

Qy   2407 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
          |||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

```
RESULT 5
ADC13230
ID   ADC13230 standard; DNA; 3930 BP.
XX
AC   ADC13230;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmut_C SEQ ID NO 9.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
```

```
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 6; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

  Query Match             99.2%;  Score 2436.4;  DB 10;  Length 3930;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-293;
  Matches 2437;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226
```

```
Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 6
ABL39960
ID   ABL39960 standard; DNA; 2463 BP.
XX
AC   ABL39960;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975YM SEQ ID NO:31.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 9; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

  Query Match             99.2%;  Score 2436.2;  DB 6;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 3.6e-293;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              ||||||||||||||||||||||||||||||   ||      |||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2455 TTC 2457
              |||
Db       2461 TTC 2463
```

RESULT 7
ADM73765
ID   ADM73765 standard; DNA; 2463 BP.
XX
AC   ADM73765;
XX
DT   03-JUN-2004 (first entry)
XX
DE   HIV-1 polynucleotide #8.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.

```
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 31; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

  Query Match             99.2%;  Score 2436.2;  DB 12;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 3.6e-293;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              ||||||||||||||||||||||||||||||   ||      |||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy         1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy         1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy         1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy         1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy         1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy         1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy         1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy         2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy         2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy         2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy         2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy         2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy         2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy         2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy         2455 TTC 2457
                |||
Db         2461 TTC 2463


RESULT 8
ADC13231
ID   ADC13231 standard; DNA; 3930 BP.
XX
AC   ADC13231;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutAtt_C SEQ ID NO 10.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 7; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;

  Query Match             99.1%;  Score 2434.8;  DB 10;  Length 3930;
  Best Local Similarity   99.9%;  Pred. No. 5.1e-293;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy           14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
                | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy           74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy          134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy          194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy          254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy          314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy          374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924


RESULT 9
ADC13232
ID   ADC13232 standard; DNA; 3930 BP.
XX
AC   ADC13232;
XX
```

```
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutIna_C SEQ ID NO 11.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 8; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

  Query Match             99.1%;  Score 2434.8;  DB 10;  Length 3930;
  Best Local Similarity   99.9%;  Pred. No. 5.1e-293;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926
```

```
Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 10
ACA03591

```
ID   ACA03591 standard; DNA; 5184 BP.
XX
AC   ACA03591;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #74.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 79; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

  Query Match             99.1%;  Score 2434.8;  DB 8;  Length 5184;
  Best Local Similarity   99.9%;  Pred. No. 4.9e-293;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980
```

```
Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 4120

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 4180

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4241 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 4300

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 4360

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 4420

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4421 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 4480

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 4540

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 4600

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 4660

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 4720

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 4780

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4781 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 4840

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4841 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 4900

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4901 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 4960

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4961 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 5020

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5021 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 5080

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5081 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 5140

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||
```

```
Db       5141 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178


RESULT 11
ADC13279
ID   ADC13279 standard; DNA; 5184 BP.
XX
AC   ADC13279;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct TatRevNefgagCpolIna_C SEQ ID NO 58.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 55; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

  Query Match             99.1%;  Score 2434.8;  DB 10;  Length 5184;
  Best Local Similarity   99.9%;  Pred. No. 4.9e-293;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 4120

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 4180

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4241 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 4300

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 4360

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 4420

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4421 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 4480

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 4540

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 4600

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 4660

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 4720

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 4780

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4781 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 4840

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4841 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 4900

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4901 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 4960

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4961 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 5020

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5021 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 5080

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5081 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 5140

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||
Db       5141 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 12
ACA03547
```
ID   ACA03547 standard; DNA; 2457 BP.
XX
AC   ACA03547;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #30.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 35; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match             98.8%;  Score 2428.6;  DB 8;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 3.2e-292;
  Matches 2447;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
```

```
Qy      121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy      181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy      241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy      301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy      361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy      421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
            |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy      481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy      541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy      601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy      661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy      721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy      781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy      841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy      901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy      961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy     1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy     1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy     1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy     1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy     1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy     1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
            ||||||||||||||||||||||||||||||   ||      |||||||||||||||||||
Db     1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy     1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy     1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy     1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy     1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy     1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy     1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy     1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy     1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy     1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy     1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy     1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy     2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy     2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy     2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy     2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy     2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 13
ADC13265
ID   ADC13265 standard; DNA; 2457 BP.
XX
AC   ADC13265;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YM_C SEQ ID NO 44.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 41; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match             98.8%;  Score 2428.6;  DB 10;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 3.2e-292;
  Matches 2447;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy         1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy         1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy         1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
                ||||||||||||||||||||||||||||   ||      |||||||||||||||||
Db         1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy         1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy         1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy         1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy         1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy         1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy         1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy         1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy         1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy         1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy         1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy         1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy         2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy         2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy         2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy         2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy         2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy         2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy         2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
RESULT 14
ABL39959
ID   ABL39959 standard; DNA; 2469 BP.
XX
AC   ABL39959;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 8; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

  Query Match             98.3%;  Score 2415.4;  DB 6;  Length 2469;
  Best Local Similarity   99.3%;  Pred. No. 1.4e-290;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368
              ||||||||||||||||||||||||||||||||||    ||      ||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
```

```
Qy       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2449 GGTGAATTC 2457
              |||||||||
Db       2461 GGTGAATTC 2469


RESULT 15
ADM73764
ID   ADM73764 standard; DNA; 2469 BP.
XX
AC   ADM73764;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #7.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 30; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

  Query Match             98.3%;  Score 2415.4;  DB 12;  Length 2469;
  Best Local Similarity   99.3%;  Pred. No. 1.4e-290;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368
              ||||||||||||||||||||||||||||||||||||   ||      |||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2449 GGTGAATTC 2457
              |||||||||
Db       2461 GGTGAATTC 2469

Search completed: June  1, 2005, 11:33:34
Job time : 1272.25 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:         June  1, 2005, 10:29:57 ; Search time 1407.9 Seconds
                                          (without alignments)
                                          10728.225 Million cell updates/sec

Title:          US-09-610-313B-32
Perfect score:  2457
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2457

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       5706582 seqs, 3073711274 residues

Total number of hits satisfying chosen parameters:      11413164

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                14:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                16:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
                17:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
                18:  /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
                19:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                20:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
                21:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                22:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2457	100.0	2457	10	US-09-899-575-32	Sequence 32, Appl
2	2443.4	99.4	2445	15	US-10-190-435-43	Sequence 43, Appl
3	2443.4	99.4	2445	16	US-10-190-305A-37	Sequence 37, Appl
4	2436.4	99.2	3930	15	US-10-190-435-9	Sequence 9, Appli
5	2436.2	99.2	2463	10	US-09-899-575-31	Sequence 31, Appl
6	2434.8	99.1	3930	15	US-10-190-435-10	Sequence 10, Appl
7	2434.8	99.1	3930	15	US-10-190-435-11	Sequence 11, Appl
8	2434.8	99.1	5184	15	US-10-190-435-58	Sequence 58, Appl
9	2434.8	99.1	5184	16	US-10-190-305A-82	Sequence 82, Appl
10	2428.6	98.8	2457	15	US-10-190-435-44	Sequence 44, Appl
11	2428.6	98.8	2457	16	US-10-190-305A-38	Sequence 38, Appl
12	2415.4	98.3	2469	10	US-09-899-575-30	Sequence 30, Appl
13	2404.4	97.9	3531	15	US-10-190-435-13	Sequence 13, Appl
14	2403.4	97.8	2457	15	US-10-190-435-45	Sequence 45, Appl
15	2403.4	97.8	2457	16	US-10-190-305A-39	Sequence 39, Appl
16	2402.8	97.8	3537	15	US-10-190-435-14	Sequence 14, Appl
17	2402.8	97.8	3537	15	US-10-190-435-15	Sequence 15, Appl
18	2401.8	97.8	5145	15	US-10-190-435-12	Sequence 12, Appl
19	2401.8	97.8	5145	16	US-10-190-305A-12	Sequence 12, Appl
20	2343.4	95.4	3607	15	US-10-190-435-48	Sequence 48, Appl
21	2343.4	95.4	3607	16	US-10-190-305A-42	Sequence 42, Appl
22	2325.2	94.6	3597	15	US-10-190-435-46	Sequence 46, Appl
23	2325.2	94.6	3597	16	US-10-190-305A-40	Sequence 40, Appl
24	2295.8	93.4	3624	15	US-10-190-435-47	Sequence 47, Appl
25	2295.8	93.4	3624	16	US-10-190-305A-41	Sequence 41, Appl
26	2136	86.9	2460	16	US-10-241-009-30	Sequence 30, Appl
27	2136	86.9	2460	16	US-10-190-434B-30	Sequence 30, Appl
28	2136	86.9	2460	16	US-10-190-305A-32	Sequence 32, Appl
29	2135	86.9	3564	16	US-10-241-009-13	Sequence 13, Appl
30	2135	86.9	3564	16	US-10-241-009-14	Sequence 14, Appl
31	2135	86.9	3564	16	US-10-190-434B-13	Sequence 13, Appl
32	2135	86.9	3564	16	US-10-190-434B-14	Sequence 14, Appl
33	2134.4	86.9	4716	15	US-10-190-435-17	Sequence 17, Appl
34	2134.4	86.9	4716	16	US-10-190-305A-13	Sequence 13, Appl
35	2130.6	86.7	3999	16	US-10-241-009-9	Sequence 9, Appli
36	2130.6	86.7	3999	16	US-10-190-434B-9	Sequence 9, Appli
37	2129	86.7	3999	16	US-10-241-009-10	Sequence 10, Appl
38	2129	86.7	3999	16	US-10-241-009-11	Sequence 11, Appl
39	2129	86.7	3999	16	US-10-190-434B-10	Sequence 10, Appl
40	2129	86.7	3999	16	US-10-190-434B-11	Sequence 11, Appl
41	2129	86.7	5283	16	US-10-241-009-54	Sequence 54, Appl
42	2129	86.7	5283	16	US-10-190-434B-54	Sequence 54, Appl
43	2129	86.7	5283	16	US-10-190-305A-75	Sequence 75, Appl
44	2128.8	86.6	4713	15	US-10-190-435-59	Sequence 59, Appl
45	2128.8	86.6	4713	16	US-10-190-305A-83	Sequence 83, Appl

ALIGNMENTS

```
RESULT 1
US-09-899-575-32
; Sequence 32, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
;  APPLICANT: Zur Megede, Jan
;  APPLICANT:  Barnett, Susan W.
;  APPLICANT:  Egnelbrecht, Susan
;  APPLICANT:  van Rensburg, Estrelita Janse
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP01631.102
;  CURRENT APPLICATION NUMBER: US/09/899,575
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: 09/475,704
;  PRIOR FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 135
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: PR975YMWM
US-09-899-575-32

  Query Match             100.0%;  Score 2457;  DB 10;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2457;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
```

```
Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340

Qy       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400

Qy       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457

RESULT 2
US-10-190-435-43
; Sequence 43, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;   LENGTH: 2445
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMWM_C
US-10-190-435-43

  Query Match             99.4%;  Score 2443.4;  DB 15;  Length 2445;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2444;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1380

Qy       1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1440

Qy       1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500

Qy       1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1501 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1560

Qy       1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1620

Qy       1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1680

Qy       1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1740

Qy       1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1800

Qy       1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1860

Qy       1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1920

Qy       1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1980

Qy       1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2040

Qy       2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2100

Qy       2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2160

Qy       2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2220

Qy       2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2280

Qy       2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2340

Qy       2347 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2400

Qy       2407 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              |||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

```
RESULT 3
US-10-190-305A-37
; Sequence 37, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
;   LENGTH: 2445
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: p2Pol.opt.YMWM_C
US-10-190-305A-37

  Query Match             99.4%;  Score 2443.4;  DB 16;  Length 2445;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2444;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840
```

```
Qy     847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy     907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy     967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy    1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy    1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy    1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy    1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy    1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy    1327 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1380

Qy    1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1440

Qy    1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500

Qy    1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1560

Qy    1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1620

Qy    1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1680

Qy    1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1740

Qy    1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1800

Qy    1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1860

Qy    1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1920

Qy    1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1980

Qy    1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2040

Qy    2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2100

Qy    2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2160

Qy    2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2220

Qy    2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2280

Qy    2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2340

Qy    2347 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2406
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2400

Qy    2407 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
           |||||||||||||||||||||||||||||||||||||||||||||
Db    2401 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 4
US-10-190-435-9
; Sequence 9, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmut_C
US-10-190-435-9

```
  Query Match             99.2%;  Score 2436.4;  DB 15;  Length 3930;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2437;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy      74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606
```

```
Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 5
US-09-899-575-31
; Sequence 31, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
; APPLICANT: Zur Megede, Jan
; APPLICANT: Barnett, Susan W.
; APPLICANT: Egnelbrecht, Susan
; APPLICANT: van Rensburg, Estrelita Janse
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP01631.102
; CURRENT APPLICATION NUMBER: US/09/899,575
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 09/475,704
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 2463
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: PR975YM
US-09-899-575-31

```
  Query Match             99.2%;  Score 2436.2;  DB 10;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              |||||||||||||||||||||||||||    ||      ||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
```

```
Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2455 TTC 2457
              |||
Db       2461 TTC 2463
```

RESULT 6
US-10-190-435-10
; Sequence 10, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt_C
US-10-190-435-10

```
  Query Match             99.1%;  Score 2434.8;  DB 15;  Length 3930;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
```

```
Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 7
US-10-190-435-11
; Sequence 11, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 3930
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna_C
US-10-190-435-11

  Query Match             99.1%;  Score 2434.8;  DB 15;  Length 3930;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546
```

```
Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 8
US-10-190-435-58
; Sequence 58, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
;   LENGTH: 5184
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: TatRevNefgagCpolIna C
US-10-190-435-58

  Query Match             99.1%;  Score 2434.8;  DB 15;  Length 5184;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 4120

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 4180

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4241 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 4300

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 4360

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 4420

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4421 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 4480

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 4540

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 4600

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 4660

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 4720

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 4780

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4781 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 4840

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4841 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 4900

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4901 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 4960

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4961 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 5020

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5021 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 5080

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5081 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 5140

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||
Db       5141 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 9
US-10-190-305A-82
; Sequence 82, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
;   LENGTH: 5184
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: TatRevNefgagCpolIna C
US-10-190-305A-82

```
  Query Match             99.1%;  Score 2434.8;  DB 16;  Length 5184;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580
```

```
Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 4120

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 4180

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4241 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 4300

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 4360

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 4420

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4421 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 4480

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 4540

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 4600

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 4660

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 4720

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 4780

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4781 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 4840

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4841 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 4900

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4901 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 4960

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4961 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 5020

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5021 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 5080

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5081 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 5140

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              ||||||||||||||||||||||||||||||||||||||
Db       5141 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

```
RESULT 10
US-10-190-435-44
; Sequence 44, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-435-44

  Query Match             98.8%;  Score 2428.6;  DB 15;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2447;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
```

```
Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              ||||||||||||||||||||||||||||||   ||      |||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy        2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy        2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 11
US-10-190-305A-38
; Sequence 38, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan
;  APPLICANT:  LIAN, Ying
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
;  TITLE OF INVENTION:  TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: 2302-18702 / 18702.002
;  CURRENT APPLICATION NUMBER: US/10/190,305A
;  CURRENT FILING DATE:  2002-07-05
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-305A-38

  Query Match             98.8%;  Score 2428.6;  DB 16;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2447;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy           1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy          61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy         121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy         181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy         241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy         301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy         361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy         421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
               |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy         481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy         541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy         601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy         661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy         721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy         781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy         841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy         901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy         961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy        1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy        1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy        1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy        1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy        1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy        1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
               ||||||||||||||||||||||||||||||   ||      |||||||||||||||||||
Db        1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy        1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy        1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy        1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy        1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
```

```
Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
RESULT 12
US-09-899-575-30
; Sequence 30, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
;  APPLICANT: Zur Megede, Jan
;  APPLICANT:  Barnett, Susan W.
;  APPLICANT:  Egnelbrecht, Susan
;  APPLICANT:  van Rensburg, Estrelita Janse
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP01631.102
;  CURRENT APPLICATION NUMBER: US/09/899,575
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: 09/475,704
;  PRIOR FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 135
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 2469
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: PR975(+)
US-09-899-575-30

  Query Match             98.3%;  Score 2415.4;  DB 10;  Length 2469;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy         901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy         961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy        1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy        1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy        1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy        1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
               ||||||||||||||||       || |  ||||||||||||||||||||||||||||||
Db        1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy        1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy        1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368
               ||||||||||||||||||||||||||||||||||    ||      |||||||||||||
Db        1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGCTACGAGCTGCACCCC 1380

Qy        1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy        1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy        1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy        1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy        1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy        1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy        1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy        1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy        1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy        1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy        1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy        2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy        2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy        2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy        2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy        2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy        2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy        2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy        2449 GGTGAATTC 2457
               |||||||||
Db        2461 GGTGAATTC 2469


RESULT 13
US-10-190-435-13
; Sequence 13, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 3531
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_C
US-10-190-435-13

  Query Match             97.9%;  Score 2404.4;  DB 15;  Length 3531;
  Best Local Similarity   99.1%;  Pred. No. 0;
  Matches 2417;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy          14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
               ||||||||| ||||||||||| || |  |    ||  ||  | || ||| |       |
Db        1088 TGGCCGAGGCGATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTTAAAA 1147

Qy          74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        1148 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1207

Qy         134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1208 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1267

Qy         194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1268 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1327

Qy         254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1328 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1387

Qy         314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1388 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1447

Qy         374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1448 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1507

Qy         434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1508 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1567

Qy         494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1568 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 1627

Qy         554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1628 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 1687

Qy         614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1688 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 1747

Qy         674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1748 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 1807

Qy         734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1808 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 1867

Qy         794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1868 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 1927

Qy         854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1928 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987

Qy         914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2047

Qy         974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2048 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2107

Qy        1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2108 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2167

Qy        1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2168 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2227

Qy        1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2228 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2287

Qy        1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2288 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2347

Qy        1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2348 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2407

Qy        1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2408 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2467

Qy        1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2468 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2527

Qy        1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2528 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2587

Qy        1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2588 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 2647

Qy        1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2648 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 2707

Qy        1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2708 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 2767

Qy        1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2768 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 2827

Qy        1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2828 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 2887

Qy        1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2888 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 2947

Qy        1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2948 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3007

Qy        1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3008 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3067

Qy        1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3068 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3127

Qy        2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3128 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3187

Qy        2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3188 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3247

Qy        2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3248 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3307

Qy        2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3308 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3367
```

```
Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db       3368 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3427

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
Db       3428 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3487

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db       3488 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3525
```

```
RESULT 14
US-10-190-435-45
; Sequence 45, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT:  BARNETT, Susan W.
;  APPLICANT:  LIAN, Ying
;  APPLICANT:  ENGELBRECHT, Susan
;  APPLICANT:  VAN RENSBURG, Estrelita J.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;  TITLE OF INVENTION:  POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;  FILE REFERENCE: PP18133.003 / 2302-18133
;  CURRENT APPLICATION NUMBER: US/10/190,435
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 319
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-435-45

  Query Match             97.8%;  Score 2403.4;  DB 15;  Length 2457;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 2439;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;
```

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
```

```
Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 15
US-10-190-305A-39
; Sequence 39, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
;   LENGTH: 2457
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-305A-39

```
  Query Match             97.8%;  Score 2403.4;  DB 16;  Length 2457;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 2439;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              |||||||||||      || |  |||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              ||||||||||||||||||||||||||||    ||      ||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

Search completed: June 2, 2005, 04:23:48
Job time : 1415.9 secs

```
OM nucleic - nucleic search, using sw model

Run on:          May 31, 2005, 23:20:00 ; Search time 1267.33 Seconds
                                          (without alignments)
                                          11504.726 Million cell updates/sec

Title:           US-09-610-313B-31
Perfect score:   2463
Sequence:        1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2463

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters:      8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_16Dec04:*
                 1:  geneseqn1980s:*
                 2:  geneseqn1990s:*
                 3:  geneseqn2000s:*
                 4:  geneseqn2001as:*
                 5:  geneseqn2001bs:*
                 6:  geneseqn2002as:*
                 7:  geneseqn2002bs:*
                 8:  geneseqn2003as:*
                 9:  geneseqn2003bs:*
                 10:  geneseqn2003cs:*
                 11:  geneseqn2003ds:*
                 12:  geneseqn2004as:*
                 13:  geneseqn2004bs:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2463	100.0	2463	6	ABL39960	Abl39960 Synthetic
2	2463	100.0	2463	12	ADM73765	Adm73765 HIV-1 pol
3	2455.4	99.7	2457	8	ACA03547	Aca03547 Synthetic
4	2455.4	99.7	2457	10	ADC13265	Adc13265 DNA of HI
5	2442.2	99.2	2469	6	ABL39959	Abl39959 Synthetic
6	2442.2	99.2	2469	12	ADM73764	Adm73764 HIV-1 pol
7	2436.2	98.9	2457	6	ABL39961	Abl39961 Synthetic
8	2436.2	98.9	2457	12	ADM73766	Adm73766 HIV-1 pol
9	2430.2	98.7	2457	8	ACA03548	Aca03548 Synthetic
10	2430.2	98.7	2457	10	ADC13266	Adc13266 DNA of HI
11	2422.6	98.4	2445	8	ACA03546	Aca03546 Synthetic
12	2422.6	98.4	2445	10	ADC13264	Adc13264 DNA of HI
13	2415.6	98.1	3930	10	ADC13230	Adc13230 DNA of HI
14	2414	98.0	3930	10	ADC13231	Adc13231 DNA of HI
15	2414	98.0	3930	10	ADC13232	Adc13232 DNA of HI
16	2414	98.0	5184	8	ACA03591	Aca03591 Synthetic
17	2414	98.0	5184	10	ADC13279	Adc13279 DNA of HI
18	2383.6	96.8	3531	10	ADC13234	Adc13234 DNA of HI
19	2382	96.7	3537	10	ADC13236	Adc13236 DNA of HI
20	2381	96.7	5145	8	ACA03521	Aca03521 Synthetic
21	2381	96.7	5145	10	ADC13233	Adc13233 DNA of HI
22	2371	96.3	3538	10	ADC13235	Adc13235 DNA of HI
23	2322.6	94.3	3607	8	ACA03551	Aca03551 Synthetic
24	2322.6	94.3	3607	10	ADC13269	Adc13269 DNA of HI
25	2322.6	94.3	3624	8	ACA03550	Aca03550 Synthetic
26	2322.6	94.3	3624	10	ADC13268	Adc13268 DNA of HI
27	2304.4	93.6	3597	8	ACA03549	Aca03549 Synthetic
28	2304.4	93.6	3597	10	ADC13267	Adc13267 DNA of HI
29	2142	87.0	2466	8	ACA03542	Aca03542 Synthetic
30	2142	87.0	2466	8	ACC78506	Acc78506 HIV p2Pol
31	2122.8	86.2	2472	8	ACA03543	Aca03543 Synthetic
32	2122.8	86.2	2472	8	ACC78507	Acc78507 HIV p2Pol
33	2115.2	85.9	2460	8	ACA03541	Aca03541 Synthetic
34	2115.2	85.9	2460	8	ACC78505	Acc78505 HIV p2Pol
35	2114.2	85.8	3564	8	ACC78488	Acc78488 HIV GagPo
36	2114.2	85.8	3564	8	ACC78489	Acc78489 HIV GagPo
37	2113.6	85.8	4716	8	ACA03522	Aca03522 Synthetic
38	2113.6	85.8	4716	10	ADC13238	Adc13238 DNA of HI
39	2109.8	85.7	3999	8	ACC78484	Acc78484 HIV GagCo
40	2108.2	85.6	3999	8	ACC78485	Acc78485 HIV GagCo
41	2108.2	85.6	3999	8	ACC78486	Acc78486 HIV GagCo
42	2108.2	85.6	5283	8	ACA03584	Aca03584 Synthetic
43	2108.2	85.6	5283	8	ACC78529	Acc78529 HIV TatRe
44	2108	85.6	4713	8	ACA03592	Aca03592 Synthetic
45	2108	85.6	4713	10	ADC13280	Adc13280 DNA of HI

ALIGNMENTS

```
RESULT 1
ABL39960
ID   ABL39960 standard; DNA; 2463 BP.
XX
AC   ABL39960;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975YM SEQ ID NO:31.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 9; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
```

```
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 2463;  DB 6;  Length 2463;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-297;
  Matches 2463;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2461 TTC 2463
              |||
Db       2461 TTC 2463
```

```
RESULT 2
ADM73765
ID   ADM73765 standard; DNA; 2463 BP.
XX
AC   ADM73765;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #8.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 31; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 2463;  DB 12;  Length 2463;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-297;
  Matches 2463;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
```

```
Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2461 TTC 2463
              |||
Db       2461 TTC 2463

RESULT 3
ACA03547
ID   ACA03547 standard; DNA; 2457 BP.
XX
AC   ACA03547;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #30.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
```

```
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 35; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match             99.7%;  Score 2455.4;  DB 8;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-296;
  Matches 2456;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
RESULT 4
ADC13265
ID   ADC13265 standard; DNA; 2457 BP.
XX
AC   ADC13265;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YM_C SEQ ID NO 44.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 41; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match             99.7%;  Score 2455.4;  DB 10;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-296;
  Matches 2456;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
RESULT 5
ABL39959
ID   ABL39959 standard; DNA; 2469 BP.
XX
AC   ABL39959;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
```

```
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 8; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

  Query Match             99.2%;  Score 2442.2;  DB 6;  Length 2469;
  Best Local Similarity   99.6%;  Pred. No. 1.5e-294;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2455 GGTGAATTC 2463
              |||||||||
Db       2461 GGTGAATTC 2469


RESULT 6
ADM73764
ID   ADM73764 standard; DNA; 2469 BP.
XX
AC   ADM73764;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #7.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 30; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

  Query Match             99.2%;  Score 2442.2;  DB 12;  Length 2469;
  Best Local Similarity   99.6%;  Pred. No. 1.5e-294;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
```

```
Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2455 GGTGAATTC 2463
              |||||||||
Db       2461 GGTGAATTC 2469
```

RESULT 7
ABL39961
ID   ABL39961 standard; DNA; 2457 BP.
XX

```
AC   ABL39961;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975YMWM SEQ ID NO:32.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 10; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

  Query Match             98.9%;  Score 2436.2;  DB 6;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 8.3e-294;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||    ||        ||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454

Qy       2461 TTC 2463
              |||
Db       2455 TTC 2457
```

```
RESULT 8
ADM73766
ID   ADM73766 standard; DNA; 2457 BP.
XX
AC   ADM73766;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #9.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 32; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

  Query Match             98.9%;  Score 2436.2;  DB 12;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 8.3e-294;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
```

```
Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||   ||      |||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454

Qy       2461 TTC 2463
              |||
Db       2455 TTC 2457


RESULT 9
ACA03548
ID   ACA03548 standard; DNA; 2457 BP.
XX
AC   ACA03548;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #31.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 36; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

  Query Match             98.7%;  Score 2430.2;  DB 8;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 4.6e-293;
  Matches 2448;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
```

```
Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              |||||||||||      || |  |||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
RESULT 10
ADC13266
ID   ADC13266 standard; DNA; 2457 BP.
XX
AC   ADC13266;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 42; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

  Query Match             98.7%;  Score 2430.2;  DB 10;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 4.6e-293;
  Matches 2448;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;
```

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||      || |  ||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 11
ACA03546
ID   ACA03546 standard; DNA; 2445 BP.
XX
AC   ACA03546;
XX
DT   22-MAY-2003 (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #29.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR) CHIRON CORP.
XX
PI   Zur Megede J, Barnett SW, Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 34; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match   98.4%; Score 2422.6; DB 8; Length 2445;
  Best Local Similarity 99.6%; Pred. No. 4e-292;
  Matches 2441; Conservative   0; Mismatches   4; Indels   6; Gaps   1;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              ||||||||||||||||||||||||    ||      ||||||||||||||||||||||||
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACC 1374

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1434

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1494

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1554

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1614

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1674

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1794

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1854

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1914

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1974

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2034

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2094

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2154

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2214

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2274

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2334

Qy       2347 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2394

Qy       2407 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

```
RESULT 12
ADC13264
ID   ADC13264 standard; DNA; 2445 BP.
XX
AC   ADC13264;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YMWM_C SEQ ID NO 43.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 40; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
```

```
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match             98.4%;  Score 2422.6;  DB 10;  Length 2445;
  Best Local Similarity   99.6%;  Pred. No. 4e-292;
  Matches 2441;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy       1327 CACCAGAAGGAGCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              |||||||||||||||||||||||   ||      |||||||||||||||||||||||||
Db       1321 CACCAGAAGGAGCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACC 1374

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1434

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1494

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1554

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1614

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1674

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1794

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1854

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1914

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1974

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2034
```

```
Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2094

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2154

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2214

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2274

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2334

Qy       2347 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2394

Qy       2407 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

```
RESULT 13
ADC13230
ID   ADC13230 standard; DNA; 3930 BP.
XX
AC   ADC13230;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmut_C SEQ ID NO 9.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 6; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

  Query Match             98.1%;  Score 2415.6;  DB 10;  Length 3930;
  Best Local Similarity   99.6%;  Pred. No. 2.8e-291;
  Matches 2434;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
```

```
Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
              |||||||||||||||||   ||      ||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              ||||||||||||||||||||||||||||||||||||||||||||
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 14
ADC13231
ID   ADC13231 standard; DNA; 3930 BP.
XX
AC   ADC13231;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutAtt_C SEQ ID NO 10.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 7; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
```

```
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;

  Query Match            98.0%;  Score 2414;  DB 10;  Length 3930;
  Best Local Similarity   99.5%;  Pred. No. 4.4e-291;
  Matches 2433;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
              ||||||||||||||||   ||      |||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              ||||||||||||||||||||||||||||||||||||||||||||
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924


RESULT 15
ADC13232
ID   ADC13232 standard; DNA; 3930 BP.
XX
AC   ADC13232;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutIna_C SEQ ID NO 11.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 8; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

  Query Match             98.0%;  Score 2414;  DB 10;  Length 3930;
  Best Local Similarity   99.5%;  Pred. No. 4.4e-291;
  Matches 2433;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
              |||||||||||||||||   ||      |||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
              ||||||||||||||||||||||||||||||||||||||||||||
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924

Search completed: June  1, 2005, 11:33:26
Job time : 1275.33 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         June  1, 2005, 03:50:55 ; Search time 10538.7 Seconds
                                          (without alignments)
                                          11324.506 Million cell updates/sec

Title:          US-09-610-313B-31
Perfect score:  2463
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2463

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters:      9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the total score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2463	100.0	2463	6	AX455915	AX455915 Sequence
2	2442.2	99.2	2469	6	AX455914	AX455914 Sequence
3	2436.2	98.9	2457	6	AX455916	AX455916 Sequence
4	2046	83.1	2306	6	BD263704	BD263704 Improved
5	2046	83.1	2306	6	CQ870574	CQ870574 Sequence
6	2046	83.1	2306	6	AR373387	AR373387 Sequence
7	2027.6	82.3	9166	6	AX427930	AX427930 Sequence
8	2025.2	82.2	2312	6	BD263706	BD263706 Improved
9	2025.2	82.2	2312	6	CQ870576	CQ870576 Sequence
10	2025.2	82.2	2312	6	AR373389	AR373389 Sequence
11	2019.2	82.0	2300	6	BD263705	BD263705 Improved
12	2019.2	82.0	2300	6	CQ870575	CQ870575 Sequence
13	2019.2	82.0	2300	6	AR373388	AR373388 Sequence
14	2019.2	82.0	9788	6	AX427936	AX427936 Sequence
15	2005.2	81.4	9169	6	AX427931	AX427931 Sequence
16	2001.8	81.3	9194	6	AX427926	AX427926 Sequence
17	2000.2	81.2	9194	6	AX427925	AX427925 Sequence
18	2000.2	81.2	12411	6	AX427927	AX427927 Sequence
19	1996.8	81.1	9785	6	AX427938	AX427938 Sequence
20	1993.4	80.9	9167	6	AX427933	AX427933 Sequence
21	1993.4	80.9	9170	6	AX427928	AX427928 Sequence
22	1993.4	80.9	9782	6	AX427935	AX427935 Sequence
23	1993.4	80.9	9783	6	AX427934	AX427934 Sequence
24	1993.4	80.9	9792	6	AX427932	AX427932 Sequence
25	1991.8	80.9	9189	6	AX427921	AX427921 Sequence
26	1988.8	80.7	4353	12	AF287353	AF287353 Synthetic
27	1972.2	80.1	3015	6	AX455946	AX455946 Sequence
28	1955.6	79.4	3009	6	AX455987	AX455987 Sequence
29	1942.6	78.9	4319	6	BD263639	BD263639 Improved
30	1942.6	78.9	4319	6	CQ870498	CQ870498 Sequence
31	1942.6	78.9	4319	6	AR373322	AR373322 Sequence
32	1932	78.4	4352	12	AF287352	AF287352 Synthetic
33	1886.6	76.6	2577	6	AX457088	AX457088 Sequence
34	1879	76.3	4343	6	AX306428	AX306428 Sequence
35	1878.8	76.3	2305	6	BD263702	BD263702 Improved
36	1878.8	76.3	2305	6	CQ870572	CQ870572 Sequence
37	1878.8	76.3	2305	6	AR373385	AR373385 Sequence
38	1875	76.1	4341	6	AX306429	AX306429 Sequence
39	1865.4	75.7	1977	6	AX455954	AX455954 Sequence
40	1857.8	75.4	1965	6	AX455952	AX455952 Sequence
41	1852.4	75.2	7897	6	AX427923	AX427923 Sequence
42	1852	75.2	2299	6	BD263703	BD263703 Improved
43	1852	75.2	2299	6	CQ870573	CQ870573 Sequence
44	1852	75.2	2299	6	AR373386	AR373386 Sequence
45	1798.8	73.0	3024	12	AF287355	AF287355 Synthetic

ALIGNMENTS

```
RESULT 1
AX455915
LOCUS       AX455915                2463 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 31 from Patent WO0204493.
ACCESSION   AX455915
VERSION     AX455915.1  GI:21714900
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 31 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1..2463
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YM"
ORIGIN

  Query Match             100.0%;  Score 2463;  DB 6;  Length 2463;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-248;
  Matches 2463;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy       241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy       301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy       361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy      1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy      1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy      2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
```

```
Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2461 TTC 2463
              |||
Db       2461 TTC 2463


RESULT 2
AX455914
LOCUS       AX455914                2469 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 30 from Patent WO0204493.
ACCESSION   AX455914
VERSION     AX455914.1  GI:21714899
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 30 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1..2469
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975(+)"
ORIGIN

  Query Match             99.2%;  Score 2442.2;  DB 6;  Length 2469;
  Best Local Similarity   99.6%;  Pred. No. 5.4e-246;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              |||||||||||||||||      || |  |||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2455 GGTGAATTC 2463
              |||||||||
Db       2461 GGTGAATTC 2469
```

```
RESULT 3
AX455916
LOCUS       AX455916              2457 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 32 from Patent WO0204493.
ACCESSION   AX455916
VERSION     AX455916.1  GI:21714901
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 32 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1..2457
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YMWM"
ORIGIN

  Query Match             98.9%;  Score 2436.2;  DB 6;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 2.3e-245;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||   ||      |||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454

Qy       2461 TTC 2463
              |||
Db       2455 TTC 2457


RESULT 4
BD263704
LOCUS       BD263704                 2306 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263704
VERSION     BD263704.1  GI:33073472
KEYWORDS    JP 2002533124-A/71.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 71 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/71
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH    Key
                      Location/Qualifiers
            FT   source          1..2306
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1..2306
```

```
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             83.1%;  Score 2046;  DB 6;  Length 2306;
  Best Local Similarity   93.6%;  Pred. No. 1.1e-204;
  Matches 2159;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db       1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db       1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db       1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db       1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db       1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db       1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db       1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db       1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db       1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db       1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db       1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db       1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db       1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db       1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
```

```
Db     2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy     2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db     2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy     2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db     2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy     2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db     2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy     2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db     2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

```
RESULT 5
CQ870574
LOCUS       CQ870574                2306 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 82 from Patent EP1433851.
ACCESSION   CQ870574
VERSION     CQ870574.1  GI:52000090
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 82 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..2306
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt.YM"
ORIGIN

  Query Match             83.1%;  Score 2046;  DB 6;  Length 2306;
  Best Local Similarity   93.6%;  Pred. No. 1.1e-204;
  Matches 2159;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;
```

```
Qy      170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db        1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy      230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db       61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy      290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db      121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy      344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db      181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy      398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db      241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy      458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db      301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy      518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db      361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy      578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db      421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy      638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db      481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy      698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db      541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy      758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db      601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy      818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db      661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy      878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db      721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy      938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db      781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy      998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db      841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy     1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db      901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy     1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db      961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy     1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db     1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy     1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db     1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy     1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db     1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy     1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db     1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy     1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db     1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy     1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db     1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy     1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db     1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440
```

```
Qy     1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db     1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy     1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db     1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy     1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db     1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy     1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db     1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy     1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db     1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy     1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db     1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy     1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db     1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy     2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db     1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy     2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db     1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy     2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db     1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy     2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db     2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy     2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db     2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy     2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db     2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy     2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db     2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy     2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db     2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

```
RESULT 6
AR373387
LOCUS       AR373387                2306 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 82 from patent US 6602705.
ACCESSION   AR373387
VERSION     AR373387.1  GI:40075490
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 82 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1..2306
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             83.1%;  Score 2046;  DB 6;  Length 2306;
  Best Local Similarity   93.6%;  Pred. No. 1.1e-204;
  Matches 2159;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;
```

```
Qy      170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db        1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy      230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db       61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy      290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db      121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy      344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db      181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy      398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db      241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy      458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db      301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy      518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db      361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy      578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db      421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy      638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db      481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy      698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db      541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy      758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db      601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy      818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db      661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy      878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db      721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy      938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db      781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy      998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db      841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900
```

```
Qy     1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db      901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy     1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db      961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy     1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db     1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy     1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db     1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy     1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db     1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy     1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db     1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy     1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db     1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy     1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db     1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy     1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db     1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy     1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db     1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy     1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db     1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy     1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db     1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy     1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db     1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy     1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db     1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy     1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db     1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy     1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db     1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy     2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db     1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy     2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db     1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy     2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db     1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy     2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db     2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy     2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db     2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy     2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db     2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy     2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db     2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy     2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db     2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

```
RESULT 7
AX427930
LOCUS       AX427930                9166 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 168 from Patent WO0232943.
ACCESSION   AX427930
VERSION     AX427930.1  GI:21538017
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 168 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9166
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match           82.3%;  Score 2027.6;  DB 6;  Length 9166;
  Best Local Similarity   91.7%;  Pred. No. 6.6e-203;
  Matches 2194;  Conservative    0;  Mismatches  179;  Indels   19;  Gaps    4;

Qy       14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db     2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy       74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db     3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy      134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db     3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy      194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db     3141 AGGACTGCACCGAGCGACAGGCTAA-TTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3199

Qy      254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db     3200 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3259

Qy      314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
```

```
Db   3260 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3319
Qy    365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db   3320 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3379
Qy    422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db   3380 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3439
Qy    482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db   3440 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3499
Qy    542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db   3500 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3559
Qy    602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db   3560 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3619
Qy    662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db   3620 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3679
Qy    722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db   3680 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3739
Qy    782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db   3740 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3799
Qy    842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db   3800 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3859
Qy    902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db   3860 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3919
Qy    962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db   3920 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3979
Qy   1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db   3980 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4039
Qy   1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db   4040 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4099
Qy   1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db   4100 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4159
Qy   1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db   4160 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4219
Qy   1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db   4220 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4279
Qy   1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1375
Db   4280 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4339
Qy   1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1435
Db   4340 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4399
Qy   1436 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1495
Db   4400 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4459
Qy   1496 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1555
Db   4460 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4519
Qy   1556 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1615
Db   4520 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4579
Qy   1616 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1675
Db   4580 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4639
Qy   1676 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1735
Db   4640 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4699
Qy   1736 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1795
Db   4700 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4759
Qy   1796 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1855
Db   4760 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4819
Qy   1856 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1915
Db   4820 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4879
Qy   1916 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1975
Db   4880 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4939
Qy   1976 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2035
Db   4940 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4999
Qy   2036 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2095
Db   5000 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5059
Qy   2096 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2155
Db   5060 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5119
Qy   2156 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2215
Db   5120 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5179
Qy   2216 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2275
Db   5180 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5239
Qy   2276 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2335
Db   5240 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5299
Qy   2336 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db   5300 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5351
```

RESULT 8
BD263706
LOCUS BD263706 2312 bp DNA linear PAT 17-JUL-2003
DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.
ACCESSION BD263706

```
VERSION     BD263706.1  GI:33073474
KEYWORDS    JP 2002533124-A/73.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 73 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/73
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH   Key
                        Location/Qualifiers
            FT   source          1..2312
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1..2312
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             82.2%;  Score 2025.2;  DB 6;  Length 2312;
  Best Local Similarity   93.3%;  Pred. No. 1.7e-202;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1351
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1411
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1412 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1471
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1531
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1532 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1592 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1651
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1652 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1711
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1712 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1771
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620
```

```
Qy       1772 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1831
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1832 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1891
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1892 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1951
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1952 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2011
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2012 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2071
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2072 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2131
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2132 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2191
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2192 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2251
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2311
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2312 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2371
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2372 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2431
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2432 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 9
CQ870576
LOCUS       CQ870576                2312 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 84 from Patent EP1433851.
ACCESSION   CQ870576
VERSION     CQ870576.1  GI:52000092
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 84 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..2312
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt(+)"
```

```
ORIGIN

  Query Match             82.2%;  Score 2025.2;  DB 6;  Length 2312;
  Best Local Similarity   93.3%;  Pred. No. 1.7e-202;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020
```

```
Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1351
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1411
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1412 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1471
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1531
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1532 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1592 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1651
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1652 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1711
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1712 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1771
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1772 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1831
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1832 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1891
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1892 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1951
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1952 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2011
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2012 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2071
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2072 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2131
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2132 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2191
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2192 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2251
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2311
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2312 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2371
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2372 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2431
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2432 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312


RESULT 10
AR373389
LOCUS       AR373389                2312 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 84 from patent US 6602705.
ACCESSION   AR373389
VERSION     AR373389.1  GI:40075492
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 84 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             82.2%;  Score 2025.2;  DB 6;  Length 2312;
  Best Local Similarity   93.3%;  Pred. No. 1.7e-202;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480
```

```
Qy    638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db    481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy    698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db    541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy    758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db    601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy    818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db    661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy    878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db    721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy    938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db    781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy    998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db    841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy   1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db    901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy   1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db    961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy   1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
Db   1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy   1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
Db   1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy   1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1351
Db   1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy   1352 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1411
Db   1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy   1412 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1471
Db   1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy   1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1531
Db   1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy   1532 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591
Db   1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy   1592 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1651
Db   1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy   1652 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1711
Db   1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy   1712 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1771
Db   1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy   1772 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1831
Db   1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy   1832 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1891
Db   1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy   1892 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1951
Db   1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy   1952 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2011
Db   1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy   2012 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2071
Db   1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy   2072 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2131
Db   1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy   2132 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2191
Db   1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy   2192 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2251
Db   2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy   2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2311
Db   2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy   2312 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2371
Db   2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy   2372 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2431
Db   2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy   2432 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2463
Db   2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 11
BD263705
LOCUS       BD263705               2300 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263705
VERSION     BD263705.1  GI:33073473
KEYWORDS    JP 2002533124-A/72.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2300)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 72 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/72
```

```
PD   08-OCT-2002
PF   30-DEC-1999 JP 2000591193
PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
KARIN HARTOG,
PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
C12N5/10,
PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
Description of Artificial Sequence: FS(-).protmod.RTopt.YMWM FH
Key          Location/Qualifiers
FT   source        1. .2300
FT                 /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2300
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match          82.0%;  Score 2019.2;  DB 6;  Length 2300;
  Best Local Similarity   93.2%;  Pred. No. 7.1e-202;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1198

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db       1199 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1254

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db       1255 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db       1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1374

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db       1375 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db       1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1494

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db       1495 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1554

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db       1555 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1614

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db       1615 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db       1675 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db       1735 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1794

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
```

```
Db       1795 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1854

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077

Db       1855 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1914

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137

Db       1915 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197

Db       1975 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2034

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257

Db       2035 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2094

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317

Db       2095 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2154

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377

Db       2155 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2214

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437

Db       2215 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2274

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463

Db       2275 TTCCCGGGGCTAGCACCGGTGAATTC 2300


RESULT 12
CQ870575
LOCUS       CQ870575                2300 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 83 from Patent EP1433851.
ACCESSION   CQ870575
VERSION     CQ870575.1  GI:52000091
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 83 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..2300
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt.YMWM"
ORIGIN

  Query Match             82.0%;  Score 2019.2;  DB 6;  Length 2300;
  Best Local Similarity   93.2%;  Pred. No. 7.1e-202;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757

Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297

Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357

Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1198

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
```

```
Db       1199 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1254

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db       1255 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db       1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1374

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db       1375 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db       1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1494

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db       1495 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1554

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db       1555 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1614

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db       1615 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db       1675 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db       1735 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1794

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db       1795 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1854

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db       1855 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1914

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db       1915 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db       1975 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2034

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db       2035 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2094

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db       2095 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2154

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db       2155 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2214

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db       2215 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2274

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2275 TTCCCGGGGCTAGCACCGGTGAATTC 2300
```

```
RESULT 13
AR373388
LOCUS       AR373388                2300 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 83 from patent US 6602705.
ACCESSION   AR373388
VERSION     AR373388.1  GI:40075491
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2300)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 83 05-AUG-2003;
FEATURES             Location/Qualifiers
     source          1..2300
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             82.0%;  Score 2019.2;  DB 6;  Length 2300;
  Best Local Similarity   93.2%;  Pred. No. 7.1e-202;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
```

```
Db     661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy     878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db     721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy     938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db     781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy     998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db     841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy    1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db     901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy    1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db     961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy    1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db    1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy    1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db    1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy    1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db    1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1198

Qy    1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db    1199 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1254

Qy    1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db    1255 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314

Qy    1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db    1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1374

Qy    1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db    1375 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434

Qy    1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db    1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1494

Qy    1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db    1495 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1554

Qy    1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db    1555 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1614

Qy    1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db    1615 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674

Qy    1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db    1675 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734

Qy    1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db    1735 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1794

Qy    1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db    1795 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1854

Qy    2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db    1855 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1914

Qy    2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db    1915 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974

Qy    2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db    1975 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2034

Qy    2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db    2035 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2094

Qy    2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db    2095 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2154

Qy    2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db    2155 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2214

Qy    2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db    2215 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2274

Qy    2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db    2275 TTCCCGGGGCTAGCACCGGTGAATTC 2300
```

```
RESULT 14
AX427936
LOCUS       AX427936                9788 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 174 from Patent WO0232943.
ACCESSION   AX427936
VERSION     AX427936.1  GI:21538023
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 174 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9788
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.0%;  Score 2019.2;  DB 6;  Length 9788;
  Best Local Similarity   91.6%;  Pred. No. 4.8e-202;
  Matches 2190;  Conservative    0;  Mismatches  178;  Indels   24;  Gaps    4;

Qy      14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db    2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy      74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
```

```
Db     3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy      134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db     3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy      194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db     3141 AGGACTGCACCGAGCGACAGGCTAA------TAGGGAAGATCTGGCCTTCCCACAAGGGA 3194

Qy      254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db     3195 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3254

Qy      314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db     3255 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3314

Qy      365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db     3315 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3374

Qy      422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db     3375 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3434

Qy      482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db     3435 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3494

Qy      542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db     3495 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3554

Qy      602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db     3555 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3614

Qy      662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db     3615 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3674

Qy      722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db     3675 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3734

Qy      782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db     3735 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3794

Qy      842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db     3795 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3854

Qy      902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db     3855 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3914

Qy      962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db     3915 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3974

Qy     1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db     3975 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4034

Qy     1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db     4035 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4094

Qy     1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db     4095 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4154

Qy     1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db     4155 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4214

Qy     1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db     4215 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4274

Qy     1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1375
Db     4275 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4334

Qy     1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1435
Db     4335 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4394

Qy     1436 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1495
Db     4395 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4454

Qy     1496 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1555
Db     4455 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4514

Qy     1556 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1615
Db     4515 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4574

Qy     1616 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1675
Db     4575 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4634

Qy     1676 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1735
Db     4635 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4694

Qy     1736 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1795
Db     4695 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4754

Qy     1796 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1855
Db     4755 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4814

Qy     1856 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1915
Db     4815 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4874

Qy     1916 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1975
Db     4875 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4934

Qy     1976 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2035
Db     4935 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4994

Qy     2036 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2095
Db     4995 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5054

Qy     2096 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2155
Db     5055 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5114

Qy     2156 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2215
Db     5115 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5174

Qy     2216 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2275
Db     5175 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5234
```

```
Qy       2276 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2335
Db       5235 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5294

Qy       2336 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db       5295 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5346


RESULT 15
AX427931
LOCUS       AX427931                9169 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 169 from Patent WO0232943.
ACCESSION   AX427931
VERSION     AX427931.1  GI:21538018
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 169 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1..9169
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             81.4%;  Score 2005.2;  DB 6;  Length 9169;
  Best Local Similarity   91.4%;  Pred. No. 1.4e-200;
  Matches 2187;  Conservative    0;  Mismatches  183;  Indels   22;  Gaps    5;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2967 TGGCCGAGGCCATGAGCCAGGTGCAGCACACCAACATCATGATGCAGCGCGGCAACTTCA 3026

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       3027 AGGGCCAGAAGCGCAT---CAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3083

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       3084 ACTGCCGCGCCCTGCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 3143

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       3144 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3202

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       3203 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3262

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db       3263 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3322

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db       3323 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3382

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db       3383 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3442

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db       3443 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3502

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db       3503 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3562

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db       3563 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3622

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db       3623 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3682

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db       3683 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3742

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db       3743 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3802

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db       3803 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3862

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db       3863 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3922

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db       3923 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3982

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db       3983 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4042

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db       4043 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4102

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db       4103 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4162

Qy       1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db       4163 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4222

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db       4223 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4282

Qy       1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1375
Db       4283 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4342

Qy       1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1435
Db       4343 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4402

Qy       1436 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1495
Db       4403 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4462

Qy       1496 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1555
Db       4463 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4522

Qy       1556 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1615
Db       4523 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4582

Qy       1616 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCAGGACCAGT 1675
```

```
Db     4583 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4642

Qy     1676 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1735
Db     4643 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4702

Qy     1736 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1795
Db     4703 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4762

Qy     1796 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1855
Db     4763 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4822

Qy     1856 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1915
Db     4823 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4882

Qy     1916 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1975
Db     4883 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4942

Qy     1976 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2035
Db     4943 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 5002

Qy     2036 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2095
Db     5003 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5062

Qy     2096 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2155
Db     5063 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5122

Qy     2156 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2215
Db     5123 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5182

Qy     2216 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2275
Db     5183 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5242

Qy     2276 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2335
Db     5243 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5302

Qy     2336 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db     5303 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5354
```

Search completed: June 1, 2005, 20:20:37
Job time : 10550.7 secs

```
                         GenCore version 5.1.6
               Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          June  1, 2005, 09:47:50 ; Search time 389.948 Seconds
                                           (without alignments)
                                           10360.282 Million cell updates/sec

Title:           US-09-610-313B-30
Perfect score:   2469
Sequence:        1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2469

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters:      2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        Issued_Patents_NA:*
                 1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                 2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                 3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                 4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                 5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                 6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the total score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                  %
Result           Query
  No.    Score   Match Length DB  ID                          Description
--------------------------------------------------------------------------------
    1     2052   83.1   2312  4  US-09-475-515-84            Sequence 84, Appl
    2   2025.2   82.0   2306  4  US-09-475-515-82            Sequence 82, Appl
    3   1998.4   80.9   2300  4  US-09-475-515-83            Sequence 83, Appl
    4   1969.4   79.8   4319  4  US-09-475-515-6             Sequence 6, Appli
    5     1858   75.3   2305  4  US-09-475-515-80            Sequence 80, Appl
    6   1831.2   74.2   2299  4  US-09-475-515-81            Sequence 81, Appl
    7   1678.6   68.0   4307  4  US-09-552-950-2             Sequence 2, Appli
    8   1659.4   67.2   4307  4  US-09-936-572-2             Sequence 2, Appli
    9   1651.4   66.9   4327  4  US-09-936-572-14            Sequence 14, Appl
   10   1651.4   66.9   4353  4  US-09-936-572-13            Sequence 13, Appl
   11   1651.4   66.9   4642  4  US-09-936-572-12            Sequence 12, Appl
   12   1651.4   66.9   9772  4  US-09-552-950-5             Sequence 5, Appli
   13   1592.8   64.5   8366  4  US-09-872-733A-6            Sequence 6, Appli
   14   1557.4   63.1   4338  4  US-09-872-733A-1            Sequence 1, Appli
   15   1336.4   54.1   2577  4  US-09-952-060-1             Sequence 1, Appli
   16   1336.4   54.1   2650  4  US-09-952-060-5             Sequence 5, Appli
   17   1327.4   53.8   4053  4  US-09-952-060-34            Sequence 34, Appl
   18   1325.2   53.7   2577  4  US-09-952-060-3             Sequence 3, Appli
   19   1325.2   53.7   2650  4  US-09-952-060-7             Sequence 7, Appli
   20   1325.2   53.7  38519  4  US-09-952-060-28            Sequence 28, Appl
   21   1227.2   49.7   9010  4  US-09-184-418C-8            Sequence 8, Appli
   22   1199.4   48.6   9913  4  US-09-827-688-11            Sequence 11, Appl
   23   1196.2   48.4   8972  4  US-09-184-418C-9            Sequence 9, Appli
   24   1190.6   48.2   2467  4  US-09-872-733A-3            Sequence 3, Appli
   25     1189   48.2   8959  4  US-09-184-418C-11           Sequence 11, Appl
   26   1165.6   47.2   8992  4  US-09-184-418C-4            Sequence 4, Appli
   27     1132   45.8   2601  3  US-09-117-217-7             Sequence 7, Appli
   28     1132   45.8   2601  3  US-09-117-217-9             Sequence 9, Appli
   29     1132   45.8   2601  3  US-09-117-217-11            Sequence 11, Appl
   30     1132   45.8   2601  3  US-09-117-217-13            Sequence 13, Appl
   31     1132   45.8   2601  4  US-09-735-487-7             Sequence 7, Appli
   32     1132   45.8   2601  4  US-09-735-487-9             Sequence 9, Appli
   33     1132   45.8   2601  4  US-09-735-487-11            Sequence 11, Appl
   34     1132   45.8   2601  4  US-09-735-487-13            Sequence 13, Appl
   35     1132   45.8   4307  4  US-09-552-950-1             Sequence 1, Appli
   36     1132   45.8   4307  4  US-09-936-572-1             Sequence 1, Appli
   37     1132   45.8   9719  4  US-09-700-304-1             Sequence 1, Appli
   38   1128.8   45.7   9050  4  US-09-184-418C-7            Sequence 7, Appli
   39   1125.6   45.6   7399  2  US-08-418-848A-9            Sequence 9, Appli
   40   1125.6   45.6   9709  2  US-08-188-583-5             Sequence 5, Appli
   41   1125.6   45.6   9709  3  US-08-388-353-1             Sequence 1, Appli
   42   1125.6   45.6   9709  3  US-08-488-551B-1            Sequence 1, Appli
   43   1125.6   45.6   9709  3  US-09-309-572-15            Sequence 15, Appl
   44   1125.6   45.6   9709  4  US-09-718-096-15            Sequence 15, Appl
   45   1125.6   45.6  12479  4  US-09-318-138-13            Sequence 13, Appl

                                ALIGNMENTS

RESULT 1
US-09-475-515-84
; Sequence 84, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
;  APPLICANT: BARNETT, Susan
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT: SRIVASTAVA, Indresh
;  APPLICANT: LIAN, Ying
;  APPLICANT: HARTOG, Karin
;  APPLICANT: LIU, Hong
;  APPLICANT: GREER, Catherine
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 84
;   LENGTH: 2312
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(-).protmod.RTopt(+)
US-09-475-515-84

  Query Match             83.1%;  Score 2052;  DB 4;  Length 2312;
  Best Local Similarity   93.6%;  Pred. No. 0;
  Matches 2165;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              |||||||||||||||  |||||||||||||||||||||  ||||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              ||||||||||| ||||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      ||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240
```

```
Qy    398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db    241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy    458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db    301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy    518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db    361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy    578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db    421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy    638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db    481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy    698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db    541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy    758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db    601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy    818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db    661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy    878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db    721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy    938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db    781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy    998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db    841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy   1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db    901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy   1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db    961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy   1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db   1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy   1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db   1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy   1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db   1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy   1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db   1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy   1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db   1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy   1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db   1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy   1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db   1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy   1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db   1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy   1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db   1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy   1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db   1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy   1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db   1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy   1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db   1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy   1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db   1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy   1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db   1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy   2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db   1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy   2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db   1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy   2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db   1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy   2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db   2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy   2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db   2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy   2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db   2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy   2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db   2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy   2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db   2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

RESULT 2
US-09-475-515-82
; Sequence 82, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:

```
;  APPLICANT: BARNETT, Susan
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT: SRIVASTAVA, Indresh
;  APPLICANT: LIAN, Ying
;  APPLICANT: HARTOG, Karin
;  APPLICANT: LIU, Hong
;  APPLICANT: GREER, Catherine
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
;   LENGTH: 2306
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(-).protmod.RTopt.YM
US-09-475-515-82

  Query Match             82.0%;  Score 2025.2;  DB 4;  Length 2306;
  Best Local Similarity   93.3%;  Pred. No. 0;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db       1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1194

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1195 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1254

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1255 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1314

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1315 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1374

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1375 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1434

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1435 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1494

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1495 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1554

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1555 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1614

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1615 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1674

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1675 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1734

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
```

```
Db      1735 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGC 1794

Qy      1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017

Db      1795 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1854

Qy      2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077

Db      1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1914

Qy      2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137

Db      1915 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1974

Qy      2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197

Db      1975 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2034

Qy      2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257

Db      2035 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2094

Qy      2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317

Db      2095 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2154

Qy      2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377

Db      2155 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2214

Qy      2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437

Db      2215 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2274

Qy      2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469

Db      2275 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2306
```

RESULT 3
US-09-475-515-83
; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt.YMWM
US-09-475-515-83

```
  Query Match             80.9%;  Score 1998.4;  DB 4;  Length 2300;
  Best Local Similarity   92.9%;  Pred. No. 0;
  Matches 2147;  Conservative    0;  Mismatches  141;  Indels   24;  Gaps    4;

Qy       170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db         1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy       230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db        61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy       290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db       121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy       344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db       181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy       398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db       241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy       458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db       301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy       518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db       361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy       578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db       421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy       638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db       481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy       698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757

Db       541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy       758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db       601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy       818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db       661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy       878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db       721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy       938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db       781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy       998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db       841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy      1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db       901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy      1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db       961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy      1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237

Db      1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074
```

```
Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db       1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGC 1194

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1195 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1248

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1249 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1308

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1309 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1368

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1369 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1428

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1429 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1488

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1489 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1548

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1549 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1608

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1609 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1668

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1669 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1728

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1729 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1788

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1789 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1848

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1849 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1908

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1909 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1968

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1969 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2028

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2029 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2088

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2089 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2148

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db       2149 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2208

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db       2209 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2268

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db       2269 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2300
```

```
RESULT 4
US-09-475-515-6
; Sequence 6, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
;  APPLICANT: BARNETT, Susan
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT: SRIVASTAVA, Indresh
;  APPLICANT: LIAN, Ying
;  APPLICANT: HARTOG, Karin
;  APPLICANT: LIU, Hong
;  APPLICANT: GREER, Catherine
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 4319
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: HIV-Gag-polymerase
US-09-475-515-6

  Query Match             79.8%;  Score 1969.4;  DB 4;  Length 4319;
  Best Local Similarity   90.1%;  Pred. No. 0;
  Matches 2158;  Conservative    0;  Mismatches  221;  Indels   16;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70
Db       1100 TGGCCGAGGCGATGAGCCAGGTGACGAACCCGGCGACCATCATGATGCAGCGCGGCAACT 1159

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db       1160 TCCGCAACCAGCGGAAGACCGTCAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCA 1219

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       1220 GGAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGCGCTGCGGCCGCGAAGGACACCAAA 1279

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       1280 TGAAAGATTGCACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAG 1338

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       1339 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 1398

Qy        311 TGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db       1399 TTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 1458

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db       1459 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 1518
```

```
Qy        419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db       1519 TCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCGCCGACGACACCGTGCTGGAGG 1578

Qy        479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db       1579 AGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGGATCGGGGGCTTCATCA 1638

Qy        539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db       1639 AGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCGGCCACAAGGCCATCGGCACCG 1698

Qy        599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
Db       1699 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 1758

Qy        659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
Db       1759 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGCCCGTGAAGCTGAAGCCGGGGA 1818

Qy        719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
Db       1819 TGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 1878

Qy        779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
Db       1879 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 1938

Qy        839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
Db       1939 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 1998

Qy        899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
Db       1999 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 2058

Qy        959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
Db       2059 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 2118

Qy       1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
Db       2119 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 2178

Qy       1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
Db       2179 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 2238

Qy       1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db       2239 GCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 2298

Qy       1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258
Db       2299 CCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 2358

Qy       1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318
Db       2359 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 2418

Qy       1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378
Db       2419 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 2478

Qy       1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438
Db       2479 CCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 2538

Qy       1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
Db       2539 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 2598

Qy       1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
Db       2599 TGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGATCCCCCTGA 2658

Qy       1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
Db       2659 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 2718

Qy       1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
Db       2719 AGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCAGGGCC 2778

Qy       1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
Db       2779 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 2838

Qy       1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
Db       2839 GCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGGTGA 2898

Qy       1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
Db       2899 GCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCATCCAGAAGG 2958

Qy       1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
Db       2959 AGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 3018

Qy       1919 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
Db       3019 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCGTGG 3078

Qy       1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
Db       3079 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 3138

Qy       2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
Db       3139 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGAGCATCGCCGACACCACCAACCAGA 3198

Qy       2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
Db       3199 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 3258

Qy       2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218
Db       3259 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 3318

Qy       2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278
Db       3319 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 3378

Qy       2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338
Db       3379 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGAGCGCCGGCA 3438

Qy       2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db       3439 TCCGCAAGGTGCTGTTCCTGAACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 3493
```

RESULT 5
US-09-475-515-80
; Sequence 80, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine

```
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
;   LENGTH: 2305
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(+).proinact.RTopt.YM
US-09-475-515-80

  Query Match             75.3%;  Score 1858;  DB 4;  Length 2305;
  Best Local Similarity   89.0%;  Pred. No. 5.3e-309;
  Matches 2058;  Conservative    0;  Mismatches  235;  Indels   19;  Gaps    4;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTT-AGGG 59

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         60 AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA 119

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        120 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG 179

Qy        344 CCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        180 CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        240 GACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAG 299

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG 359

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        360 GGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTG 419

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        480 ATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGC 539

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        540 CCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        600 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 659

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        660 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 719

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        720 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 779

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        780 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 839

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        840 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 899

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        900 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 959

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        960 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1019

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db       1020 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1073

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db       1074 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1133

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1134 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1193

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1194 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1253

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1254 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1313

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1314 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1373

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1374 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1433

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1434 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1493

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1494 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1553

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1554 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1613

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1614 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1673

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1674 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1733

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1734 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1793

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1794 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1853
```

```
Qy     2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db     1854 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1913

Qy     2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db     1914 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1973

Qy     2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db     1974 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2033

Qy     2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db     2034 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2093

Qy     2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db     2094 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2153

Qy     2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db     2154 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2213

Qy     2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db     2214 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2273

Qy     2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db     2274 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2305
```

RESULT 6
US-09-475-515-81
; Sequence 81, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
; LENGTH: 2299
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(+).proinact.RTopt.YMWM
US-09-475-515-81

Query Match 74.2%; Score 1831.2; DB 4; Length 2299;
Best Local Similarity 88.6%; Pred. No. 2e-304;
Matches 2049; Conservative 0; Mismatches 238; Indels 25; Gaps 5;

```
Qy      170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db        1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTT-AGGG 59

Qy      230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db       60 AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA 119

Qy      290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db      120 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG 179

Qy      344 CCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGC 397
Db      180 CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239

Qy      398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db      240 GACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAG 299

Qy      458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db      300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG 359

Qy      518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db      360 GGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTG 419

Qy      578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db      420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479

Qy      638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db      480 ATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGC 539

Qy      698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db      540 CCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599

Qy      758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db      600 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 659

Qy      818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db      660 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 719

Qy      878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db      720 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 779

Qy      938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db      780 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 839

Qy      998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db      840 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 899

Qy     1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db      900 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 959

Qy     1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db      960 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1019

Qy     1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db     1020 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1073

Qy     1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db     1074 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1133

Qy     1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db     1134 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGC 1193
```

```
Qy   1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db   1194 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1247

Qy   1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db   1248 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1307

Qy   1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db   1308 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1367

Qy   1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db   1368 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1427

Qy   1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db   1428 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1487

Qy   1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db   1488 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1547

Qy   1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db   1548 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1607

Qy   1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db   1608 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1667

Qy   1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db   1668 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1727

Qy   1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db   1728 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1787

Qy   1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db   1788 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1847

Qy   2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db   1848 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1907

Qy   2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db   1908 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1967

Qy   2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db   1968 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2027

Qy   2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db   2028 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2087

Qy   2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db   2088 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2147

Qy   2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db   2148 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2207

Qy   2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db   2208 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2267

Qy   2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db   2268 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2299

RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
;  APPLICANT: Oxford Biomedica (UK) Limited
;  TITLE OF INVENTION: Anti-Viral Vectors
;  FILE REFERENCE: 674524-2004
;  CURRENT APPLICATION NUMBER: US/09/552,950
;  CURRENT FILING DATE:  2000-04-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4307
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:gagpol-SYNgp - codon
;   OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

  Query Match             68.0%;  Score 1678.6;  DB 4;  Length 4307;
  Best Local Similarity   82.9%;  Pred. No. 2.8e-278;
  Matches 1968;  Conservative    0;  Mismatches  389;  Indels   16;  Gaps    4;

Qy     12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db   1086 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1145

Qy     69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db   1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1205

Qy    129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db   1206 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGCTGGAAATGCGGCAAGGAAGGCCACCA 1265

Qy    189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db   1266 GATGAAAGACTGTACGGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACA 1324

Qy    249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db   1325 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCCGCGCCAACAGCCCCACCCGCCGCGA 1384

Qy    309 GCTGCAGGT------GCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db   1385 GCTTCAGGTCTGGGGTCGCGACAACAACTCCCCCTCCGAAGCAGGAGCCGACCGCCAGGG 1444

Qy    363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db   1445 CACGGTGTCCTTCAACTTCCCTCAGGTCACGCTTTGGCAGCGACCCCTCGTCACCATCAA 1504

Qy    417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db   1505 GATCGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1564

Qy    477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db   1565 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1624

Qy    537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db   1625 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1684

Qy    597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db   1685 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 1744
```

```
Qy     657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db    1745 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 1804

Qy     717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db    1805 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 1864

Qy     777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db    1865 GGAGATTTGCACAGAGATGGAAAAGGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCC 1924

Qy     837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db    1925 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 1984

Qy     897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db    1985 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2044

Qy     957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db    2045 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2104

Qy    1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db    2105 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2164

Qy    1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db    2165 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2224

Qy    1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db    2225 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2284

Qy    1197 CCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGAT 1256
Db    2285 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2344

Qy    1257 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1316
Db    2345 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2404

Qy    1317 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCA 1376
Db    2405 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 2464

Qy    1377 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1436
Db    2465 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2524

Qy    1437 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1496
Db    2525 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2584

Qy    1497 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1556
Db    2585 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2644

Qy    1557 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1616
Db    2645 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 2704

Qy    1617 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1676
Db    2705 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 2764

Qy    1677 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1736
Db    2765 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 2824

Qy    1737 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1796
Db    2825 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 2884

Qy    1797 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1856
Db    2885 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 2944

Qy    1857 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1916
Db    2945 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3004

Qy    1917 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1976
Db    3005 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3064

Qy    1977 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2036
Db    3065 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3124

Qy    2037 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2096
Db    3125 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3184

Qy    2097 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2156
Db    3185 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3244

Qy    2157 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2216
Db    3245 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3304

Qy    2217 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2276
Db    3305 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3364

Qy    2277 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2336
Db    3365 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3424

Qy    2337 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2369
Db    3425 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3457
```

RESULT 8
US-09-936-572-2
; Sequence 2, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gagpol-SYNgp-codon optimised gagpol sequence
US-09-936-572-2

Query Match 67.2%; Score 1659.4; DB 4; Length 4307;
Best Local Similarity 82.4%; Pred. No. 5.3e-275;

```
Matches 1956;  Conservative    0;  Mismatches  401;  Indels   16;  Gaps    4;

Qy        12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db      1086 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1145

Qy        69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db      1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1205

Qy       129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db      1206 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGCTGGAAATGCGGCAAGGAAGGCCACCA 1265

Qy       189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db      1266 GATGAAAGACTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACA 1324

Qy       249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db      1325 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGA 1384

Qy       309 GCTGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db      1385 GCTTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGG 1444

Qy       363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db      1445 AACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATAAA 1504

Qy       417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db      1505 GATAGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1564

Qy       477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db      1565 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1624

Qy       537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db      1625 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1684

Qy       597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db      1685 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 1744

Qy       657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db      1745 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 1804

Qy       717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db      1805 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 1864

Qy       777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db      1865 GGAGATTTGCACAGAGATGGAAAAGGAAGGAAAAATCTCCAAGATTGGGCCTGAGAACCC 1924

Qy       837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db      1925 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 1984

Qy       897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db      1985 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2044

Qy       957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db      2045 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2104

Qy      1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db      2105 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2164

Qy      1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db      2165 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2224

Qy      1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db      2225 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2284

Qy      1197 CCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGAT 1256
Db      2285 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2344

Qy      1257 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1316
Db      2345 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2404

Qy      1317 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCA 1376
Db      2405 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 2464

Qy      1377 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1436
Db      2465 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2524

Qy      1437 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1496
Db      2525 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2584

Qy      1497 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1556
Db      2585 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2644

Qy      1557 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1616
Db      2645 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 2704

Qy      1617 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1676
Db      2705 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 2764

Qy      1677 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1736
Db      2765 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 2824

Qy      1737 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1796
Db      2825 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 2884

Qy      1797 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1856
Db      2885 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 2944

Qy      1857 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1916
Db      2945 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3004

Qy      1917 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1976
Db      3005 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3064

Qy      1977 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2036
Db      3065 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3124

Qy      2037 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2096
Db      3125 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3184

Qy      2097 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2156
Db      3185 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3244
```

```
Qy       2157 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2216
Db       3245 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3304

Qy       2217 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2276
Db       3305 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3364

Qy       2277 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2336
Db       3365 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3424

Qy       2337 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2369
Db       3425 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3457
```

RESULT 9
US-09-936-572-14
; Sequence 14, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 4327
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: pSYNGP4-codon
;   OTHER INFORMATION: optimised HIV-1 gagpol with 20bp of the leader
;   OTHER INFORMATION: sequence of HIV-1
US-09-936-572-14

```
  Query Match             66.9%;  Score 1651.4;  DB 4;  Length 4327;
  Best Local Similarity   82.2%;  Pred. No. 1.2e-273;
  Matches 1951;  Conservative    0;  Mismatches  406;  Indels   16;  Gaps    4;

Qy         12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db       1106 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1165

Qy         69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db       1166 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1225

Qy        129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db       1226 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCA 1285

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db       1286 AATGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACA 1344

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db       1345 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGA 1404

Qy        309 GCTGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db       1405 GCTTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGG 1464

Qy        363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db       1465 AACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAA 1524

Qy        417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db       1525 GATAGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1584

Qy        477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db       1585 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1644

Qy        537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db       1645 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1704

Qy        597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db       1705 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 1764

Qy        657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db       1765 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 1824

Qy        717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db       1825 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 1884

Qy        777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db       1885 GGAGATTTGCACAGAGATGGAAAAGGAAGGAAAAATCTCCAAGATTGGGCCTGAGAACCC 1944

Qy        837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db       1945 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 2004

Qy        897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db       2005 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2064

Qy        957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db       2065 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2124

Qy       1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db       2125 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2184

Qy       1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db       2185 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2244

Qy       1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db       2245 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2304

Qy       1197 CCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGAT 1256
Db       2305 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2364

Qy       1257 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1316
Db       2365 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2424

Qy       1317 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCA 1376
Db       2425 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 2484

Qy       1377 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1436
Db       2485 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2544

Qy       1437 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1496
```

```
Db   2545 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2604

Qy   1497 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1556
Db   2605 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2664

Qy   1557 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1616
Db   2665 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 2724

Qy   1617 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1676
Db   2725 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 2784

Qy   1677 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1736
Db   2785 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 2844

Qy   1737 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1796
Db   2845 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 2904

Qy   1797 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1856
Db   2905 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 2964

Qy   1857 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1916
Db   2965 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3024

Qy   1917 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1976
Db   3025 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3084

Qy   1977 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2036
Db   3085 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3144

Qy   2037 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2096
Db   3145 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3204

Qy   2097 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2156
Db   3205 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3264

Qy   2157 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2216
Db   3265 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3324

Qy   2217 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2276
Db   3325 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3384

Qy   2277 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2336
Db   3385 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3444

Qy   2337 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2369
Db   3445 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3477
```

RESULT 10
US-09-936-572-13
; Sequence 13, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 4353
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP3-codon
; OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence from
; OTHER INFORMATION: the major splice donor
US-09-936-572-13

```
  Query Match             66.9%;  Score 1651.4;  DB 4;  Length 4353;
  Best Local Similarity   82.2%;  Pred. No. 1.2e-273;
  Matches 1951;  Conservative    0;  Mismatches  406;  Indels   16;  Gaps    4;

Qy     12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db   1132 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1191

Qy     69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db   1192 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1251

Qy    129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db   1252 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCA 1311

Qy    189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db   1312 AATGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACA 1370

Qy    249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db   1371 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGA 1430

Qy    309 GCTGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db   1431 GCTTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGG 1490

Qy    363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db   1491 AACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAA 1550

Qy    417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db   1551 GATAGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1610

Qy    477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db   1611 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1670

Qy    537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db   1671 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1730

Qy    597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db   1731 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 1790

Qy    657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db   1791 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 1850

Qy    717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db   1851 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 1910
```

```
Qy        777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db       1911 GGAGATTTGCACAGAGATGGAAAAGGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCC 1970

Qy        837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db       1971 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 2030

Qy        897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db       2031 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2090

Qy        957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db       2091 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2150

Qy       1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db       2151 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2210

Qy       1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db       2211 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2270

Qy       1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db       2271 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2330

Qy       1197 CCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGAT 1256
Db       2331 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2390

Qy       1257 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1316
Db       2391 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2450

Qy       1317 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCA 1376
Db       2451 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 2510

Qy       1377 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1436
Db       2511 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2570

Qy       1437 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1496
Db       2571 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2630

Qy       1497 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1556
Db       2631 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2690

Qy       1557 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1616
Db       2691 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 2750

Qy       1617 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1676
Db       2751 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 2810

Qy       1677 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1736
Db       2811 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 2870

Qy       1737 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1796
Db       2871 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 2930

Qy       1797 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1856
Db       2931 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 2990

Qy       1857 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1916
Db       2991 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3050

Qy       1917 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1976
Db       3051 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3110

Qy       1977 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2036
Db       3111 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3170

Qy       2037 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2096
Db       3171 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3230

Qy       2097 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2156
Db       3231 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3290

Qy       2157 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2216
Db       3291 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3350

Qy       2217 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2276
Db       3351 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3410

Qy       2277 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2336
Db       3411 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3470

Qy       2337 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2369
Db       3471 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3503
```

RESULT 11
US-09-936-572-12
; Sequence 12, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 4642
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: pSYNGP2-codon
;   OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence
US-09-936-572-12

```
  Query Match             66.9%;  Score 1651.4;  DB 4;  Length 4642;
  Best Local Similarity   82.2%;  Pred. No. 1.2e-273;
  Matches 1951;  Conservative    0;  Mismatches  406;  Indels   16;  Gaps    4;

Qy         12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db       1421 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1480

Qy         69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db       1481 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1540
```

```
Qy     129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db    1541 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCA 1600

Qy     189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db    1601 AATGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTAGGGAAGATCTGGCCTTCCCACA 1659

Qy     249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db    1660 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGA 1719

Qy     309 GCTGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db    1720 GCTTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGG 1779

Qy     363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db    1780 AACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAA 1839

Qy     417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db    1840 GATAGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1899

Qy     477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db    1900 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1959

Qy     537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db    1960 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 2019

Qy     597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db    2020 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 2079

Qy     657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db    2080 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 2139

Qy     717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db    2140 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 2199

Qy     777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db    2200 GGAGATTTGCACAGAGATGGAAAAGGAAGGAAAAATCTCCAAGATTGGGCCTGAGAACCC 2259

Qy     837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db    2260 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 2319

Qy     897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db    2320 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2379

Qy     957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db    2380 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2439

Qy    1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db    2440 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2499

Qy    1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db    2500 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2559

Qy    1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db    2560 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2619

Qy    1197 CCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGAT 1256
Db    2620 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2679

Qy    1257 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1316
Db    2680 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2739

Qy    1317 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCA 1376
Db    2740 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 2799

Qy    1377 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1436
Db    2800 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2859

Qy    1437 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1496
Db    2860 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2919

Qy    1497 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1556
Db    2920 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2979

Qy    1557 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1616
Db    2980 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 3039

Qy    1617 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1676
Db    3040 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 3099

Qy    1677 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1736
Db    3100 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 3159

Qy    1737 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1796
Db    3160 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 3219

Qy    1797 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1856
Db    3220 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 3279

Qy    1857 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1916
Db    3280 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3339

Qy    1917 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1976
Db    3340 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3399

Qy    1977 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2036
Db    3400 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3459

Qy    2037 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2096
Db    3460 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3519

Qy    2097 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2156
Db    3520 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3579

Qy    2157 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2216
Db    3580 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3639

Qy    2217 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2276
Db    3640 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3699

Qy    2277 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2336
```

```
Db       3700 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3759

Qy       2337 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2369

Db       3760 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3792

RESULT 12
US-09-552-950-5
; Sequence 5, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
;  APPLICANT: Oxford Biomedica (UK) Limited
;  TITLE OF INVENTION: Anti-Viral Vectors
;  FILE REFERENCE: 674524-2004
;  CURRENT APPLICATION NUMBER: US/09/552,950
;  CURRENT FILING DATE:  2000-04-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 9772
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:pSYNGP
US-09-552-950-5

  Query Match            66.9%;  Score 1651.4;  DB 4;  Length 9772;
  Best Local Similarity  82.2%;  Pred. No. 1.4e-273;
  Matches 1951;  Conservative    0;  Mismatches  406;  Indels   16;  Gaps    4;

Qy         12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68

Db       2193 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 2252

Qy         69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128

Db       2253 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 2312

Qy        129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188

Db       2313 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCA 2372

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248

Db       2373 AATGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACA 2431

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308

Db       2432 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGA 2491

Qy        309 GCTGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362

Db       2492 GCTTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGG 2551

Qy        363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416

Db       2552 AACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAA 2611

Qy        417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476

Db       2612 GATAGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 2671

Qy        477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536

Db       2672 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 2731

Qy        537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596

Db       2732 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 2791

Qy        597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656

Db       2792 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 2851

Qy        657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716

Db       2852 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 2911

Qy        717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776

Db       2912 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 2971

Qy        777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836

Db       2972 GGAGATTTGCACAGAGATGGAAAAGGAAGGAAAATCTCCAAGATTGGGCCTGAGAACCC 3031

Qy        837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896

Db       3032 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 3091

Qy        897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956

Db       3092 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 3151

Qy        957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016

Db       3152 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 3211

Qy       1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076

Db       3212 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 3271

Qy       1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136

Db       3272 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 3331

Qy       1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196

Db       3332 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 3391

Qy       1197 CCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGAT 1256

Db       3392 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 3451

Qy       1257 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1316

Db       3452 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 3511

Qy       1317 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCA 1376

Db       3512 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 3571

Qy       1377 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1436

Db       3572 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 3631

Qy       1437 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1496

Db       3632 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 3691

Qy       1497 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1556

Db       3692 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 3751

Qy       1557 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1616

Db       3752 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 3811

Qy       1617 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1676

Db       3812 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 3871

Qy       1677 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1736
```

```
Db     3872 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 3931

Qy     1737 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1796
Db     3932 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 3991

Qy     1797 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1856
Db     3992 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 4051

Qy     1857 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1916
Db     4052 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 4111

Qy     1917 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1976
Db     4112 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 4171

Qy     1977 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2036
Db     4172 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 4231

Qy     2037 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2096
Db     4232 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 4291

Qy     2097 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2156
Db     4292 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 4351

Qy     2157 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2216
Db     4352 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 4411

Qy     2217 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2276
Db     4412 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 4471

Qy     2277 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2336
Db     4472 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 4531

Qy     2337 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2369
Db     4532 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 4564
```

RESULT 13
US-09-872-733A-6
; Sequence 6, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 8366
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: DNA sequence
;   OTHER INFORMATION: of the construct pCMVgagpolBNKan containing a CMV
;   OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin
;   OTHER INFORMATION: resistance gene

US-09-872-733A-6

```
  Query Match             64.5%;  Score 1592.8;  DB 4;  Length 8366;
  Best Local Similarity   80.7%;  Pred. No. 1.4e-263;
  Matches 1914;  Conservative    0;  Mismatches  442;  Indels   16;  Gaps    4;

Qy       14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70
Db     1857 TGGCCGAGGCGATGAGCCAGGTGACGAACTCGGCGACCATAATGATGCAGAGAGGCAACT 1916

Qy       71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db     1917 TCCGGAACCAGCGGAAGATCGTCAAGTGCTTCAATTGTGGCAAAGAAGGGCACACCGCCA 1976

Qy      131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db     1977 GGAACTGCCGGGCCCCCCGGAAGAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAA 2036

Qy      191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db     2037 TGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAG 2095

Qy      251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db     2096 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 2155

Qy      311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db     2156 TTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAA 2215

Qy      365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db     2216 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 2275

Qy      419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db     2276 TCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAG 2335

Qy      479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db     2336 AAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGGATCGGGGGCTTCATCA 2395

Qy      539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db     2396 AGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 2455

Qy      599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
Db     2456 TATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATCGGCT 2515

Qy      659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
Db     2516 GCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCGGGGA 2575

Qy      719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
Db     2576 TGGACGGCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCTTAGTCG 2635

Qy      779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
Db     2636 AAATCTGTACAGAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCT 2695

Qy      839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
Db     2696 ACAACACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGG 2755

Qy      899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
Db     2756 ACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCAC 2815

Qy      959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
Db     2816 ATCCCGCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACT 2875
```

```
Qy       1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
Db       2876 TCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCA 2935

Qy       1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
Db       2936 ACAACGAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGAT 2995

Qy       1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db       2996 CACCAGCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACC 3055

Qy       1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258
Db       3056 CAGACATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCG 3115

Qy       1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318
Db       3116 GGCAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCA 3175

Qy       1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378
Db       3176 CACCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCATC 3235

Qy       1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438
Db       3236 CTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACG 3295

Qy       1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
Db       3296 ACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAG 3355

Qy       1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
Db       3356 TTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCCACTGA 3415

Qy       1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
Db       3416 CAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGTACATG 3475

Qy       1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
Db       3476 GAGTGTACTACGACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCAAGGCC 3535

Qy       1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
Db       3536 AATGGACCTACCAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTACGCAA 3595

Qy       1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
Db       3596 GGATGAGGGGTGCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAAGATCA 3655

Qy       1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
Db       3656 CCACAGAGAGCATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACAGAAGG 3715

Qy       1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
Db       3716 AGACATGGGAGACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGGAGT 3775

Qy       1919 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
Db       3776 TCGTGAACACCCCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGG 3835

Qy       1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
Db       3836 GAGCAGAGACCTTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAAGGCAG 3895

Qy       2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
Db       3896 GCTACGTGACCAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAACCAGA 3955

Qy       2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
Db       3956 AGACTGAGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCG 4015

Qy       2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218
Db       4016 TGACAGACTCACAGTACGCACTGGGCATCATCCAAGCACAACCAGACCAATCCGAGTCAG 4075

Qy       2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278
Db       4076 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGCATGGG 4135

Qy       2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338
Db       4136 TACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGGA 4195

Qy       2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db       4196 TCCGGAAGGTGCTGTTCCTGGACGGATCGAT 4227
```

RESULT 14
US-09-872-733A-1
; Sequence 1, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4338
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
;   OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene
US-09-872-733A-1

```
  Query Match             63.1%;  Score 1557.4;  DB 4;  Length 4338;
  Best Local Similarity   80.5%;  Pred. No. 1.5e-257;
  Matches 1914;  Conservative    0;  Mismatches  441;  Indels   22;  Gaps    7;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70
Db       1085 TGGCCGAGGCGATGAGCCAGGTGACGAACTCGGCGACCATAATGATGCAGAGAGGCAACT 1144

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db       1145 TCCGGAACCAGCGGAAGATCGTCAAGTGCTTCAATTGTGGCAAAGAAGGGCACACCGCCA 1204

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       1205 GGAACTGCCGGGCCCCCCGGAAGAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAA 1264

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       1265 TGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAG 1323

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       1324 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 1383

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
```

```
Db       1384 TTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAA 1443

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418

Db       1444 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 1503

Qy        419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478

Db       1504 TCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAG 1563

Qy        479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538

Db       1564 AAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGGATCGGGGGCTTCATCA 1623

Qy        539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598

Db       1624 AGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 1683

Qy        599 TGCTGATCGGCCCCACC----CCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654

Db       1684 TATTAGTAGGACCTACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATC 1743

Qy        655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714

Db       1744 GGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCG 1803

Qy        715 GGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTG 774

Db       1804 GGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCTTA 1863

Qy        775 ACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAAC 834

Db       1864 GTCGAAATCTGTACAGAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAAC 1923

Qy        835 CCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTG 894

Db       1924 CCCTACAACACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTG 1983

Qy        895 GTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954

Db       1984 GTGGACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATC 2043

Qy        955 CCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCC 1014

Db       2044 CCACATCCCGCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCC 2103

Qy       1015 TACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGC 1074

Db       2104 TACTTCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGC 2163

Qy       1075 ATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAG 1134

Db       2164 ATCAACAACGAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAG 2223

Qy       1135 GGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGC 1194

Db       2224 GGATCACCAGCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAA 2283

Qy       1195 AACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAG 1254

Db       2284 AACCCAGACATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAG 2343

Qy       1255 ATC-GGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTT 1313

Db       2344 ATCGGGGCAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACT 2403

Qy       1314 CACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCT 1373

Db       2404 GACCACACCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACT 2463

Qy       1374 GCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGT 1433

Db       2464 GCATCCTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGT 2523

Qy       1434 GAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCAT 1493

Db       2524 GAACGACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCAT 2583

Qy       1494 CAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCC 1553

Db       2584 CAAAGTTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCC 2643

Qy       1554 CCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGT 1613

Db       2644 ACTGACAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGT 2703

Qy       1614 GCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCA 1673

Db       2704 ACATGGAGTGTACTACGACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCA 2763

Qy       1674 CGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTA 1733

Db       2764 AGGCCAATGGACCTACCAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTA 2823

Qy       1734 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1793

Db       2824 CGCAAGGATGAGGGGTGCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAA 2883

Qy       1794 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1853

Db       2884 GATCACCACAGAGAGCATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACA 2943

Qy       1854 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1913

Db       2944 GAAGGAGACATGGGAGACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTG 3003

Qy       1914 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1973

Db       3004 GGAGTTCGTGAACACCCCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCAT 3063

Qy       1974 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2033

Db       3064 CGTGGGAGCAGAGACCTTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAA 3123

Qy       2034 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2093

Db       3124 GGCAGGCTACGTGACCAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAA 3183

Qy       2094 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2153

Db       3184 CCAGAAGACTGAGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAA 3243

Qy       2154 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2213

Db       3244 CATCGTGACAGACTCACAGTACG-CATGGGCATCATCCAAGCACAACCAGACCAATCCGA 3302

Qy       2214 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2273

Db       3303 GTCAGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGC 3362

Qy       2274 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2333

Db       3363 ATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGC 3422

Qy       2334 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370

Db       3423 TGGGATCCGGAAGGTGCTGTTCCTGGACGGGATCGAT 3459
```

RESULT 15
US-09-952-060-1
; Sequence 1, Application US/09952060
; Patent No. 6733993
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.
; APPLICANT: Youil, Rima

```
; APPLICANT:  Bett, Andrew J.
; APPLICANT:  Chen, Ling
; APPLICANT:  Kaslow, David C.
; APPLICANT:  Shiver, John W.
; APPLICANT:  Toner, Timothy J.
; APPLICANT:  Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION:  VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION:  MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE:  2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2577
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Codon optimized DNA encoding modified wt pol
US-09-952-060-1

  Query Match             54.1%;  Score 1336.4;  DB 4;  Length 2577;
  Best Local Similarity   86.7%;  Pred. No. 8.3e-220;
  Matches 1472;  Conservative    0;  Mismatches  226;  Indels    0;  Gaps    0;

Qy        672 CCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAA 731
Db         15 CCCCATCTCCCCCATTGAGACTGTGCCTGTGAAGCTGAAGCCTGGCATGGATGGCCCCAA 74

Qy        732 GGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGA 791
Db         75 GGTGAAGCAGTGGCCCCTGACTGAGGAGAAGATCAAGGCCCTGGTGGAAATCTGCACTGA 134

Qy        792 GATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGT 851
Db        135 GATGGAGAAGGAGGGCAAAATCTCCAAGATTGGCCCCGAGAACCCCTACAACACCCCTGT 194

Qy        852 GTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCT 911
Db        195 GTTTGCCATCAAGAAGAAGGACTCCACCAAGTGGAGGAAGCTGGTGGACTTCAGGGAGCT 254

Qy        912 GAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCT 971
Db        255 GAACAAGAGGACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCTGGCCT 314

Qy        972 GAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCT 1031
Db        315 GAAGAAGAAGAAGTCTGTGACTGTGCTGGATGTGGGGGATGCCTACTTCTCTGTGCCCCT 374

Qy       1032 GGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCC 1091
Db        375 GGATGAGGACTTCAGGAAGTACACTGCCTTCACCATCCCCTCCATCAACAATGAGACCCC 434

Qy       1092 CGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTT 1151
Db        435 TGGCATCAGGTACCAGTACAATGTGCTGCCCCAGGGCTGGAAGGGCTCCCCTGCCATCTT 494

Qy       1152 CCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGAT 1211
Db        495 CCAGTCCTCCATGACCAAGATCCTGGAGCCCTTCAGGAAGCAGAACCCTGACATTGTGAT 554

Qy       1212 CTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGC 1271
Db        555 CTACCAGTACATGGATGACCTGTATGTGGGCTCTGACCTGGAGATTGGGCAGCACAGGAC 614

Qy       1272 CAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAA 1331
Db        615 CAAGATTGAGGAGCTGAGGCAGCACCTGCTGAGGTGGGGCCTGACCACCCCTGACAAGAA 674

Qy       1332 GCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGAC 1391
Db        675 GCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTATGAGCTGCACCCCGACAAGTGGAC 734

Qy       1392 CGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCT 1451
Db        735 TGTGCAGCCCATTGTGCTGCCTGAGAAGGACTCCTGGACTGTGAATGACATCCAGAAGCT 794

Qy       1452 GGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTG 1511
Db        795 GGTGGGCAAGCTGAACTGGGCCTCCCAAATCTACCCTGGCATCAAGGTGAGGCAGCTGTG 854

Qy       1512 CAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGA 1571
Db        855 CAAGCTGCTGAGGGGCACCAAGGCCCTGACTGAGGTGATCCCCCTGACTGAGGAGGCTGA 914

Qy       1572 GCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGA 1631
Db        915 GCTGGAGCTGGCTGAGAACAGGGAGATCCTGAAGGAGCCTGTGCATGGGGTGTACTATGA 974

Qy       1632 CCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCA 1691
Db        975 CCCCTCCAAGGACCTGATTGCTGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCA 1034

Qy       1692 GATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGC 1751
Db       1035 AATCTACCAGGAGCCCTTCAAGAACCTGAAGACTGGCAAGTATGCCAGGATGAGGGGGGC 1094

Qy       1752 CCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCAT 1811
Db       1095 CCACACCAATGATGTGAAGCAGCTGACTGAGGCTGTGCAGAAGATCACCACTGAGTCCAT 1154

Qy       1812 CGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGAC 1871
Db       1155 TGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGACCTGGGAGAC 1214

Qy       1872 CTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCC 1931
Db       1215 CTGGTGGACTGAGTACTGGCAGGCCACCTGGATCCCTGAGTGGGAGTTTGTGAACACCCC 1274

Qy       1932 CCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTT 1991
Db       1275 CCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATTGTGGGGGCTGAGACCTT 1334

Qy       1992 CTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGA 2051
Db       1335 CTATGTGGATGGGGCTGCCAACAGGGAGACCAAGCTGGGCAAGGCTGGCTATGTGACCAA 1394

Qy       2052 CCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCA 2111
Db       1395 CAGGGGCAGGCAGAAGGTGGTGACCCTGACTGACACCACCAACCAGAAGACTGAGCTCCA 1454

Qy       2112 GGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCA 2171
Db       1455 GGCCATCTACCTGGCCCTCCAGGACTCTGGCCTGGAGGTGAACATTGTGACTGACTCCCA 1514

Qy       2172 GTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCA 2231
Db       1515 GTATGCCCTGGGCATCATCCAGGCCCAGCCTGATCAGTCTGAGTCTGAGCTGGTGAACCA 1574

Qy       2232 GATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAA 2291
Db       1575 GATCATTGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGCCTGCCCACAA 1634

Qy       2292 GGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCT 2351
Db       1635 GGGCATTGGGGGCAATGAGCAGGTGGACAAGCTGGTGTCTGCTGGCATCAGGAAGGTGCT 1694
```

```
Qy       2352 GTTCCTGGACGGCATCGA 2369
              ||||||||| ||||| ||
Db       1695 GTTCCTGGATGGCATTGA 1712

Search completed: June  2, 2005, 03:12:53
Job time : 401.948 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:          June  1, 2005, 09:47:50 ; Search time 389 Seconds
                                           (without alignments)
                                           10360.282 Million cell updates/sec

Title:           US-09-610-313B-31
Perfect score:   2463
Sequence:        1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2463

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters:      2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                 1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                 2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                 3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                 4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                 5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                 6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2046	83.1	2306	4	US-09-475-515-82	Sequence 82, Appl
2	2025.2	82.2	2312	4	US-09-475-515-84	Sequence 84, Appl
3	2019.2	82.0	2300	4	US-09-475-515-83	Sequence 83, Appl
4	1942.6	78.9	4319	4	US-09-475-515-6	Sequence 6, Appli
5	1878.8	76.3	2305	4	US-09-475-515-80	Sequence 80, Appl
6	1852	75.2	2299	4	US-09-475-515-81	Sequence 81, Appl
7	1651.8	67.1	4307	4	US-09-552-950-2	Sequence 2, Appli
8	1632.6	66.3	4307	4	US-09-936-572-2	Sequence 2, Appli
9	1624.6	66.0	4327	4	US-09-936-572-14	Sequence 14, Appl
10	1624.6	66.0	4353	4	US-09-936-572-13	Sequence 13, Appl
11	1624.6	66.0	4642	4	US-09-936-572-12	Sequence 12, Appl
12	1624.6	66.0	9772	4	US-09-552-950-5	Sequence 5, Appli
13	1566	63.6	8366	4	US-09-872-733A-6	Sequence 6, Appli
14	1530.6	62.1	4338	4	US-09-872-733A-1	Sequence 1, Appli
15	1309.6	53.2	2577	4	US-09-952-060-1	Sequence 1, Appli
16	1309.6	53.2	2650	4	US-09-952-060-5	Sequence 5, Appli
17	1307	53.1	4053	4	US-09-952-060-34	Sequence 34, Appl
18	1304.8	53.0	2577	4	US-09-952-060-3	Sequence 3, Appli
19	1304.8	53.0	2650	4	US-09-952-060-7	Sequence 7, Appli
20	1304.8	53.0	38519	4	US-09-952-060-28	Sequence 28, Appl
21	1203.6	48.9	9010	4	US-09-184-418C-8	Sequence 8, Appli
22	1175.8	47.7	9913	4	US-09-827-688-11	Sequence 11, Appl
23	1172.6	47.6	8972	4	US-09-184-418C-9	Sequence 9, Appli
24	1165.4	47.3	8959	4	US-09-184-418C-11	Sequence 11, Appl
25	1163.8	47.3	2467	4	US-09-872-733A-3	Sequence 3, Appli
26	1142	46.4	8992	4	US-09-184-418C-4	Sequence 4, Appli
27	1106.8	44.9	2601	3	US-09-117-217-7	Sequence 7, Appli
28	1106.8	44.9	2601	3	US-09-117-217-9	Sequence 9, Appli
29	1106.8	44.9	2601	3	US-09-117-217-11	Sequence 11, Appl
30	1106.8	44.9	2601	3	US-09-117-217-13	Sequence 13, Appl
31	1106.8	44.9	2601	4	US-09-735-487-7	Sequence 7, Appli
32	1106.8	44.9	2601	4	US-09-735-487-9	Sequence 9, Appli
33	1106.8	44.9	2601	4	US-09-735-487-11	Sequence 11, Appl
34	1106.8	44.9	2601	4	US-09-735-487-13	Sequence 13, Appl
35	1106.8	44.9	4307	4	US-09-552-950-1	Sequence 1, Appli
36	1106.8	44.9	4307	4	US-09-936-572-1	Sequence 1, Appli
37	1106.8	44.9	9719	4	US-09-700-304-1	Sequence 1, Appli
38	1105.2	44.9	9050	4	US-09-184-418C-7	Sequence 7, Appli
39	1100.4	44.7	7399	2	US-08-418-848A-9	Sequence 9, Appli
40	1100.4	44.7	9709	2	US-08-188-583-5	Sequence 5, Appli
41	1100.4	44.7	9709	3	US-08-388-353-1	Sequence 1, Appli
42	1100.4	44.7	9709	3	US-08-488-551B-1	Sequence 1, Appli
43	1100.4	44.7	9709	3	US-09-309-572-15	Sequence 15, Appl
44	1100.4	44.7	9709	4	US-09-718-096-15	Sequence 15, Appl
45	1100.4	44.7	12479	4	US-09-318-138-13	Sequence 13, Appl

ALIGNMENTS

```
RESULT 1
US-09-475-515-82
; Sequence 82, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
;  APPLICANT: BARNETT, Susan
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT: SRIVASTAVA, Indresh
;  APPLICANT: LIAN, Ying
;  APPLICANT: HARTOG, Karin
;  APPLICANT: LIU, Hong
;  APPLICANT: GREER, Catherine
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
;   LENGTH: 2306
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(-).protmod.RTopt.YM
US-09-475-515-82

  Query Match             83.1%;  Score 2046;  DB 4;  Length 2306;
  Best Local Similarity   93.6%;  Pred. No. 0;
  Matches 2159;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              |||||||||||||||  |||||||||||||||||||||  ||||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCGCAGCGAGG 343
              |||||||||| ||||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      ||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240
```

```
Qy    398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db    241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy    458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db    301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy    518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db    361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy    578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db    421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy    638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db    481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy    698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db    541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy    758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db    601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy    818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db    661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy    878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db    721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy    938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db    781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy    998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db    841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy   1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db    901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy   1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db    961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy   1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db   1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy   1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db   1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy   1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db   1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy   1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db   1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy   1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db   1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy   1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db   1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy   1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db   1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy   1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db   1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy   1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db   1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy   1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db   1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy   1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db   1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy   1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db   1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy   1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db   1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy   1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db   1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy   2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db   1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy   2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db   1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy   2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db   1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy   2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db   2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy   2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db   2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy   2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db   2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy   2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db   2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy   2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db   2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

RESULT 2
US-09-475-515-84
; Sequence 84, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:

```
;  APPLICANT: BARNETT, Susan
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT: SRIVASTAVA, Indresh
;  APPLICANT: LIAN, Ying
;  APPLICANT: HARTOG, Karin
;  APPLICANT: LIU, Hong
;  APPLICANT: GREER, Catherine
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 84
;   LENGTH: 2312
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(-).protmod.RTopt(+)
US-09-475-515-84

  Query Match             82.2%;  Score 2025.2;  DB 4;  Length 2312;
  Best Local Similarity   93.3%;  Pred. No. 0;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1351
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1411
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1412 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1471
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1531
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1532 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1592 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1651
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1652 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1711
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1712 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1771
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1772 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1831
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1832 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1891
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1892 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1951
```

```
Db    1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy    1952 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2011
Db    1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy    2012 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2071
Db    1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy    2072 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2131
Db    1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy    2132 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2191
Db    1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy    2192 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2251
Db    2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy    2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2311
Db    2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy    2312 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2371
Db    2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy    2372 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2431
Db    2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy    2432 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2463
Db    2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 3
US-09-475-515-83
; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
;  APPLICANT: BARNETT, Susan
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT: SRIVASTAVA, Indresh
;  APPLICANT: LIAN, Ying
;  APPLICANT: HARTOG, Karin
;  APPLICANT: LIU, Hong
;  APPLICANT: GREER, Catherine
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
;   LENGTH: 2300
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(-).protmod.RTopt.YMWM
US-09-475-515-83

  Query Match             82.0%;  Score 2019.2;  DB 4;  Length 2300;
  Best Local Similarity   93.2%;  Pred. No. 0;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;
```

```
Qy     170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db       1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy     230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db      61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy     290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db     121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy     344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db     181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy     398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db     241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy     458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db     301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy     518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db     361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy     578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db     421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy     638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db     481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy     698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db     541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy     758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db     601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy     818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db     661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy     878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db     721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy     938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db     781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy     998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db     841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy    1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db     901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy    1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db     961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy    1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db    1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080
```

```
Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1198

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db       1199 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1254

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db       1255 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db       1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1374

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db       1375 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db       1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1494

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db       1495 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1554

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db       1555 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1614

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db       1615 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db       1675 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db       1735 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1794

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db       1795 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1854

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db       1855 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1914

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db       1915 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db       1975 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2034

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db       2035 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2094

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db       2095 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2154

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db       2155 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2214

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db       2215 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2274

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2275 TTCCCGGGGCTAGCACCGGTGAATTC 2300
```

```
RESULT 4
US-09-475-515-6
; Sequence 6, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
;  APPLICANT: BARNETT, Susan
;  APPLICANT: ZUR MEGEDE, Jan
;  APPLICANT: SRIVASTAVA, Indresh
;  APPLICANT: LIAN, Ying
;  APPLICANT: HARTOG, Karin
;  APPLICANT: LIU, Hong
;  APPLICANT: GREER, Catherine
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 4319
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: HIV-Gag-polymerase
US-09-475-515-6

  Query Match             78.9%;  Score 1942.6;  DB 4;  Length 4319;
  Best Local Similarity   89.7%;  Pred. No. 0;
  Matches 2129;  Conservative    0;  Mismatches  225;  Indels   19;  Gaps    4;

Qy         33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db       1122 GACGAACCCGGCGACCATCATGATGCAGCGCGGCAACTTCCGCAACCAGCGGAAGACCGT 1181

Qy         93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db       1182 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCAGGAACTGCCGCGCCCCCCGCAA 1241

Qy        153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db       1242 GAAGGGCTGCTGGCGCTGCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACA 1301

Qy        213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db       1302 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTC 1360

Qy        273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------A 326
Db       1361 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAA 1420

Qy        327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
Db       1421 CAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCA 1480

Qy        381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db       1481 GATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGATCGGCGGCCAGCTCAAGGAGGC 1540
```

```
Qy        441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db       1541 GCTGCTCGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTG 1600

Qy        501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db       1601 GAAGCCCAAGATGATCGGCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGAT 1660

Qy        561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db       1661 CCCCGTGGAGATCTGCGGCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGT 1720

Qy        621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db       1721 GAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAG 1780

Qy        681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db       1781 CCCCATCGAGACGGTGCCCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCA 1840

Qy        741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db       1841 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAA 1900

Qy        801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db       1901 GGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 1960

Qy        861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db       1961 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 2020

Qy        921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db       2021 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 2080

Qy        981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db       2081 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGA 2140

Qy       1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db       2141 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 2200

Qy       1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db       2201 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAG 2260

Qy       1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db       2261 CATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTA 2320

Qy       1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db       2321 CATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGA 2380

Qy       1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db       2381 GGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 2440

Qy       1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db       2441 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 2500

Qy       1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db       2501 CATCATGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 2560

Qy       1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
Db       2561 GCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCT 2620

Qy       1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db       2621 GCGCGGCACCAAGGCCCTGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCT 2680

Qy       1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db       2681 GGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAA 2740

Qy       1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db       2741 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCA 2800

Qy       1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db       2801 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAA 2860

Qy       1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db       2861 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTG 2920

Qy       1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db       2921 GGGCAAGATCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGAT 2980

Qy       1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db       2981 GGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 3040

Qy       1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db       3041 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGA 3100

Qy       1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db       3101 CGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGCGGCCG 3160

Qy       2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db       3161 CCAGAAGGTGGTGAGCATCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 3220

Qy       2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db       3221 CCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 3280

Qy       2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db       3281 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGA 3340

Qy       2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db       3341 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGG 3400

Qy       2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db       3401 CGGCAACGAGCAGGTGGACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAA 3460

Qy       2355 CGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db       3461 CGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 3493
```

RESULT 5
US-09-475-515-80
; Sequence 80, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine

```
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
;   LENGTH: 2305
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(+).proinact.RTopt.YM
US-09-475-515-80

  Query Match             76.3%;  Score 1878.8;  DB 4;  Length 2305;
  Best Local Similarity   89.4%;  Pred. No. 7.2e-312;
  Matches 2061;  Conservative    0;  Mismatches  232;  Indels   13;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTT-AGGG 59

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         60 AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA 119

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        120 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG 179

Qy        344 CCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        180 CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        240 GACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAG 299

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG 359

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        360 GGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTG 419

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        480 ATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGC 539

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        540 CCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        600 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 659

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        660 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 719

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        720 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 779

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        780 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 839

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        840 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 899

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        900 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 959

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        960 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1019

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1020 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1079

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1080 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1139

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db       1140 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1199

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db       1200 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1259

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db       1260 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1319

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db       1320 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1379

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db       1380 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1439

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db       1440 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1499

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db       1500 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1559

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db       1560 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1619

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db       1620 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1679

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db       1680 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1739

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db       1740 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1799

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db       1800 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1859
```

```
Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db       1860 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1919

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db       1920 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1979

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db       1980 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2039

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db       2040 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2099

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db       2100 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2159

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db       2160 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2219

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db       2220 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2279

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2280 TTCCCGGGGCTAGCACCGGTGAATTC 2305
```

RESULT 6
US-09-475-515-81
; Sequence 81, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
;   LENGTH: 2299
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(+).proinact.RTopt.YMWM
US-09-475-515-81

```
  Query Match             75.2%;  Score 1852;  DB 4;  Length 2299;
  Best Local Similarity   89.0%;  Pred. No. 2.7e-307;
  Matches 2052;  Conservative    0;  Mismatches  235;  Indels   19;  Gaps    4;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTT-AGGG 59

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         60 AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA 119

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        120 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG 179

Qy        344 CCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        180 CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        240 GACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAG 299

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG 359

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        360 GGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTG 419

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        480 ATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGC 539

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        540 CCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        600 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 659

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        660 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 719

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        720 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 779

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        780 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 839

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        840 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 899

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        900 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 959

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        960 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1019

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1020 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1079

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1080 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1139

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db       1140 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1197
```

```
Qy   1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db   1198 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1253

Qy   1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db   1254 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1313

Qy   1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db   1314 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1373

Qy   1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db   1374 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1433

Qy   1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db   1434 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1493

Qy   1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db   1494 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1553

Qy   1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db   1554 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1613

Qy   1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db   1614 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1673

Qy   1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db   1674 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1733

Qy   1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db   1734 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1793

Qy   1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db   1794 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1853

Qy   2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db   1854 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1913

Qy   2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db   1914 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1973

Qy   2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db   1974 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2033

Qy   2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db   2034 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2093

Qy   2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db   2094 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2153

Qy   2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db   2154 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2213

Qy   2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db   2214 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2273

Qy   2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db   2274 TTCCCGGGGCTAGCACCGGTGAATTC 2299
```

RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:gagpol-SYNgp - codon
; OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

```
  Query Match             67.1%;  Score 1651.8;  DB 4;  Length 4307;
  Best Local Similarity   82.5%;  Pred. No. 4.2e-273;
  Matches 1939;  Conservative    0;  Mismatches  391;  Indels   19;  Gaps    4;

Qy     33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db   1110 GACCAACTCCGCTACCATCATGATGCAGCGCGGCAACTTTCGGAACCAACGCAAGATCGT 1169

Qy     93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db   1170 CAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGCCCGCAACTGCAGGGCCCCTAGGAA 1229

Qy    153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db   1230 AAAGGGCTGCTGGAAATGCGGCAAGGAAGGCCACCAGATGAAAGACTGTACGGAGAGACA 1289

Qy    213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db   1290 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTC 1348

Qy    273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGT------GCGCGGCGA 326
Db   1349 AGAGCAGACCCGCGCCAACAGCCCCACCCGCCGCGAGCTTCAGGTCTGGGGTCGCGACAA 1408

Qy    327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
Db   1409 CAACTCCCCCTCCGAAGCAGGAGCCGACCGCCAGGGCACGGTGTCCTTCAACTTCCCTCA 1468

Qy    381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db   1469 GGTCACGCTTTGGCAGCGACCCCTCGTCACCATCAAGATCGGGGGGCAGCTCAAGGAGGC 1528

Qy    441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db   1529 TCTCCTGGACACCGGAGCAGACGACACCGTGCTGGAGGAGATGTCGTTGCCAGGCCGCTG 1588

Qy    501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db   1589 GAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCATCAAGGTGCGCCAGTATGACCAGAT 1648

Qy    561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db   1649 CCTCATCGAAATCTGCGGCCACAAGGCTATCGGTACCGTGCTGGTGGGCCCCACACCCGT 1708

Qy    621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db   1709 CAACATCATCGGACGCAACCTGTTGACGCAGATCGGTTGCACGCTGAACTTCCCCATTAG 1768
```

```
Qy        681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db       1769 CCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGGGATGGACGGCCCGAAGGTCAAGCA 1828

Qy        741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db       1829 ATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGTGGAGATTTGCACAGAGATGGAAAA 1888

Qy        801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db       1889 GGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCCGTACAACACGCCGGTGTTCGCAAT 1948

Qy        861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db       1949 CAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 2008

Qy        921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db       2009 CACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCCGCACCCCGCAGGGCTGAAGAAGAA 2068

Qy        981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db       2069 GAAATCCGTGACCGTACTGGATGTGGGTGATGCCTACTTCTCCGTTCCCCTGGACGAAGA 2128

Qy       1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db       2129 CTTCAGGAAGTACACTGCCTTCACAATCCCTTCGATCAACAACGAGACACCGGGGATTCG 2188

Qy       1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db       2189 ATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGGCTCTCCCGCAATCTTCCAGAGTAG 2248

Qy       1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db       2249 CATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAACCCCGACATCGTCATCTATCAGTA 2308

Qy       1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db       2309 CATGGATGACTTGTACGTGGGCTCTGATCTAGAGATAGGGCAGCACCGCACCAAGATCGA 2368

Qy       1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db       2369 GGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGACCACACCCGACAAGAAGCACCAGAA 2428

Qy       1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db       2429 GGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCACCCTGACAAATGGACCGTGCAGCC 2488

Qy       1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db       2489 TATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAACGACATACAGAAGCTGGTGGGGAA 2548

Qy       1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
Db       2549 GTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAAGGTGAGGCAGCTGTGCAAACTCCT 2608

Qy       1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db       2609 CCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCTAACCGAGGAGGCCGAGCTCGAACT 2668

Qy       1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db       2669 GGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCACGGCGTGTACTATGACCCCTCCAA 2728

Qy       1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db       2729 GGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGGCCAGTGGACCTATCAGATTTACCA 2788

Qy       1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db       2789 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGGATGAGGGGTGCCCACACTAA 2848

Qy       1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db       2849 CGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGATCACCACCGAAAGCATCGTGATCTG 2908

Qy       1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db       2909 GGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAAGGAAACCTGGGAAACCTGGTGGAC 2968

Qy       1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db       2969 AGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGAGTTCGTCAACACCCCTCCCCTGGT 3028

Qy       1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db       3029 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGTGGGCGCCGAAACCTTCTACGTGGA 3088

Qy       1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db       3089 TGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGCCGGATACGTCACTAACCGGGGCAG 3148

Qy       2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db       3149 ACAGAAGGTTGTCACCCTCACTGACACCACCAACCAGAAGACTGAGCTGCAGGCCATTTA 3208

Qy       2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db       3209 CCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACATCGTGACAGACTCTCAGTATGCCCT 3268

Qy       2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db       3269 GGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTCCGAGCTGGTCAATCAGATCATCGA 3328

Qy       2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db       3329 GCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTGGGTACCCGCCCACAAAGGCATTGG 3388

Qy       2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db       3389 CGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGGCATCAGGAAGGTGCTATTCCTGGA 3448

Qy       2355 CGGCATCGA 2363
Db       3449 TGGCATCGA 3457
```

RESULT 8
US-09-936-572-2
; Sequence 2, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4307
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: gagpol-SYNgp-codon optimised gagpol sequence
US-09-936-572-2

  Query Match   66.3%; Score 1632.6; DB 4; Length 4307;
  Best Local Similarity   82.0%; Pred. No. 8e-270;

```
Matches 1927;  Conservative    0;  Mismatches  403;  Indels   19;  Gaps    4;

Qy         33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db       1110 GACCAACTCCGCTACCATCATGATGCAGCGCGGCAACTTTCGGAACCAACGCAAGATCGT 1169

Qy         93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db       1170 CAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGCCCGCAACTGCAGGGCCCCTAGGAA 1229

Qy        153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db       1230 AAAGGGCTGCTGGAAATGCGGCAAGGAAGGCCACCAGATGAAAGACTGTACTGAGAGACA 1289

Qy        213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db       1290 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTC 1348

Qy        273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGG------CGA 326
Db       1349 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAA 1408

Qy        327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
Db       1409 CAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCA 1468

Qy        381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db       1469 GATCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGGGCAGCTCAAGGAGGC 1528

Qy        441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db       1529 TCTCCTGGACACCGGAGCAGACGACACCGTGCTGGAGGAGATGTCGTTGCCAGGCCGCTG 1588

Qy        501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db       1589 GAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCATCAAGGTGCGCCAGTATGACCAGAT 1648

Qy        561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db       1649 CCTCATCGAAATCTGCGGCCACAAGGCTATCGGTACCGTGCTGGTGGGCCCCACACCCGT 1708

Qy        621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db       1709 CAACATCATCGGACGCAACCTGTTGACGCAGATCGGTTGCACGCTGAACTTCCCCATTAG 1768

Qy        681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db       1769 CCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGGGATGGACGGCCCGAAGGTCAAGCA 1828

Qy        741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db       1829 ATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGTGGAGATTTGCACAGAGATGGAAAA 1888

Qy        801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db       1889 GGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCCGTACAACACGCCGGTGTTCGCAAT 1948

Qy        861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db       1949 CAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 2008

Qy        921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db       2009 CACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCCGCACCCCGCAGGGCTGAAGAAGAA 2068

Qy        981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db       2069 GAAATCCGTGACCGTACTGGATGTGGGTGATGCCTACTTCTCCGTTCCCCTGGACGAAGA 2128

Qy       1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db       2129 CTTCAGGAAGTACACTGCCTTCACAATCCCTTCGATCAACAACGAGACACCGGGGATTCG 2188

Qy       1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db       2189 ATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGGCTCTCCCGCAATCTTCCAGAGTAG 2248

Qy       1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db       2249 CATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAACCCCGACATCGTCATCTATCAGTA 2308

Qy       1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db       2309 CATGGATGACTTGTACGTGGGCTCTGATCTAGAGATAGGGCAGCACCGCACCAAGATCGA 2368

Qy       1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db       2369 GGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGACCACACCCGACAAGAAGCACCAGAA 2428

Qy       1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db       2429 GGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCACCCTGACAAATGGACCGTGCAGCC 2488

Qy       1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db       2489 TATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAACGACATACAGAAGCTGGTGGGGAA 2548

Qy       1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
Db       2549 GTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAAGGTGAGGCAGCTGTGCAAACTCCT 2608

Qy       1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db       2609 CCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCTAACCGAGGAGGCCGAGCTCGAACT 2668

Qy       1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db       2669 GGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCACGGCGTGTACTATGACCCCTCCAA 2728

Qy       1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db       2729 GGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGGCCAGTGGACCTATCAGATTTACCA 2788

Qy       1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db       2789 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGGATGAGGGGTGCCCACACTAA 2848

Qy       1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db       2849 CGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGATCACCACCGAAAGCATCGTGATCTG 2908

Qy       1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db       2909 GGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAAGGAAACCTGGGAAACCTGGTGGAC 2968

Qy       1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db       2969 AGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGAGTTCGTCAACACCCCTCCCCTGGT 3028

Qy       1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db       3029 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGTGGGCGCCGAAACCTTCTACGTGGA 3088

Qy       1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db       3089 TGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGCCGGATACGTCACTAACCGGGGCAG 3148

Qy       2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db       3149 ACAGAAGGTTGTCACCCTCACTGACACCACCAACCAGAAGACTGAGCTGCAGGCCATTTA 3208

Qy       2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db       3209 CCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACATCGTGACAGACTCTCAGTATGCCCT 3268
```

```
Qy       2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db       3269 GGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTCCGAGCTGGTCAATCAGATCATCGA 3328

Qy       2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db       3329 GCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTGGGTACCCGCCCACAAAGGCATTGG 3388

Qy       2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db       3389 CGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGGCATCAGGAAGGTGCTATTCCTGGA 3448

Qy       2355 CGGCATCGA 2363
Db       3449 TGGCATCGA 3457
```

RESULT 9
US-09-936-572-14
; Sequence 14, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 4327
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP4-codon
; OTHER INFORMATION: optimised HIV-1 gagpol with 20bp of the leader
; OTHER INFORMATION: sequence of HIV-1
US-09-936-572-14

```
  Query Match             66.0%;  Score 1624.6;  DB 4;  Length 4327;
  Best Local Similarity   81.8%;  Pred. No. 1.9e-268;
  Matches 1922;  Conservative    0;  Mismatches  408;  Indels   19;  Gaps    4;

Qy         33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db       1130 GACCAACTCCGCTACCATCATGATGCAGCGCGGCAACTTTCGGAACCAACGCAAGATCGT 1189

Qy         93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db       1190 CAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGCCCGCAACTGCAGGGCCCCTAGGAA 1249

Qy        153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db       1250 AAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACA 1309

Qy        213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db       1310 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCTTC 1368

Qy        273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGG------CGA 326
Db       1369 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAAGAGACAA 1428

Qy        327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
Db       1429 CAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAGCTTCCCTCA 1488

Qy        381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db       1489 GATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAGGGGGGCAGCTCAAGGAGGC 1548

Qy        441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db       1549 TCTCCTGGACACCGGAGCAGACGACACCGTGCTGGAGGAGATGTCGTTGCCAGGCCGCTG 1608

Qy        501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db       1609 GAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCATCAAGGTGCGCCAGTATGACCAGAT 1668

Qy        561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db       1669 CCTCATCGAAATCTGCGGCCACAAGGCTATCGGTACCGTGCTGGTGGGCCCCACACCCGT 1728

Qy        621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db       1729 CAACATCATCGGACGCAACCTGTTGACGCAGATCGGTTGCACGCTGAACTTCCCCATTAG 1788

Qy        681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db       1789 CCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGGGATGGACGGCCCGAAGGTCAAGCA 1848

Qy        741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db       1849 ATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGTGGAGATTTGCACAGAGATGGAAAA 1908

Qy        801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db       1909 GGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCCGTACAACACGCCGGTGTTCGCAAT 1968

Qy        861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db       1969 CAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 2028

Qy        921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db       2029 CACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCCGCACCCCGCAGGGCTGAAGAAGAA 2088

Qy        981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db       2089 GAAATCCGTGACCGTACTGGATGTGGGTGATGCCTACTTCTCCGTTCCCCTGGACGAAGA 2148

Qy       1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db       2149 CTTCAGGAAGTACACTGCCTTCACAATCCCTTCGATCAACAACGAGACACCGGGGATTCG 2208

Qy       1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db       2209 ATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGGCTCTCCCGCAATCTTCCAGAGTAG 2268

Qy       1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db       2269 CATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAACCCCGACATCGTCATCTATCAGTA 2328

Qy       1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db       2329 CATGGATGACTTGTACGTGGGCTCTGATCTAGAGATAGGGCAGCACCGCACCAAGATCGA 2388

Qy       1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db       2389 GGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGACCACACCCGACAAGAAGCACCAGAA 2448

Qy       1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db       2449 GGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCACCCTGACAAATGGACCGTGCAGCC 2508

Qy       1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db       2509 TATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAACGACATACAGAAGCTGGTGGGGAA 2568

Qy       1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
```

```
Db   2569 GTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAAGGTGAGGCAGCTGTGCAAACTCCT 2628

Qy   1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db   2629 CCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCTAACCGAGGAGGCCGAGCTCGAACT 2688

Qy   1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db   2689 GGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCACGGCGTGTACTATGACCCCTCCAA 2748

Qy   1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db   2749 GGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGGCCAGTGGACCTATCAGATTTACCA 2808

Qy   1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db   2809 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGGATGAGGGGTGCCCACACTAA 2868

Qy   1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db   2869 CGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGATCACCACCGAAAGCATCGTGATCTG 2928

Qy   1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db   2929 GGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAAGGAAACCTGGGAAACCTGGTGGAC 2988

Qy   1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db   2989 AGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGAGTTCGTCAACACCCCTCCCCTGGT 3048

Qy   1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db   3049 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGTGGGCGCCGAAACCTTCTACGTGGA 3108

Qy   1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db   3109 TGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGCCGGATACGTCACTAACCGGGGCAG 3168

Qy   2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db   3169 ACAGAAGGTTGTCACCCTCACTGACACCACCAACCAGAAGACTGAGCTGCAGGCCATTTA 3228

Qy   2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db   3229 CCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACATCGTGACAGACTCTCAGTATGCCCT 3288

Qy   2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db   3289 GGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTCCGAGCTGGTCAATCAGATCATCGA 3348

Qy   2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db   3349 GCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTGGGTACCCGCCCACAAAGGCATTGG 3408

Qy   2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db   3409 CGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGGCATCAGGAAGGTGCTATTCCTGGA 3468

Qy   2355 CGGCATCGA 2363
Db   3469 TGGCATCGA 3477
```

RESULT 10
US-09-936-572-13
; Sequence 13, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 4353
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP3-codon
; OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence from
; OTHER INFORMATION: the major splice donor
US-09-936-572-13

```
Query Match             66.0%;  Score 1624.6;  DB 4;  Length 4353;
Best Local Similarity   81.8%;  Pred. No. 1.9e-268;
Matches 1922;  Conservative    0;  Mismatches  408;  Indels   19;  Gaps    4;

Qy     33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db   1156 GACCAACTCCGCTACCATCATGATGCAGCGCGGCAACTTTCGGAACCAACGCAAGATCGT 1215

Qy     93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db   1216 CAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGCCCGCAACTGCAGGGCCCCTAGGAA 1275

Qy    153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db   1276 AAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACA 1335

Qy    213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db   1336 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCTTC 1394

Qy    273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGG------CGA 326
Db   1395 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAAGAGACAA 1454

Qy    327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
Db   1455 CAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAGCTTCCCTCA 1514

Qy    381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db   1515 GATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAGGGGGGCAGCTCAAGGAGGC 1574

Qy    441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db   1575 TCTCCTGGACACCGGAGCAGACGACACCGTGCTGGAGGAGATGTCGTTGCCAGGCCGCTG 1634

Qy    501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db   1635 GAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCATCAAGGTGCGCCAGTATGACCAGAT 1694

Qy    561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db   1695 CCTCATCGAAATCTGCGGCCACAAGGCTATCGGTACCGTGCTGGTGGGCCCCACACCCGT 1754

Qy    621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db   1755 CAACATCATCGGACGCAACCTGTTGACGCAGATCGGTTGCACGCTGAACTTCCCCATTAG 1814

Qy    681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db   1815 CCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGGGATGGACGGCCCGAAGGTCAAGCA 1874

Qy    741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db   1875 ATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGTGGAGATTTGCACAGAGATGGAAAA 1934
```

```
Qy        801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db       1935 GGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCCGTACAACACGCCGGTGTTCGCAAT 1994

Qy        861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db       1995 CAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 2054

Qy        921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db       2055 CACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCCGCACCCCGCAGGGCTGAAGAAGAA 2114

Qy        981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db       2115 GAAATCCGTGACCGTACTGGATGTGGGTGATGCCTACTTCTCCGTTCCCCTGGACGAAGA 2174

Qy       1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db       2175 CTTCAGGAAGTACACTGCCTTCACAATCCCTTCGATCAACAACGAGACACCGGGGATTCG 2234

Qy       1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db       2235 ATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGGCTCTCCCGCAATCTTCCAGAGTAG 2294

Qy       1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db       2295 CATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAACCCCGACATCGTCATCTATCAGTA 2354

Qy       1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db       2355 CATGGATGACTTGTACGTGGGCTCTGATCTAGAGATAGGGCAGCACCGCACCAAGATCGA 2414

Qy       1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db       2415 GGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGACCACACCCGACAAGAAGCACCAGAA 2474

Qy       1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db       2475 GGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCACCCTGACAAATGGACCGTGCAGCC 2534

Qy       1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db       2535 TATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAACGACATACAGAAGCTGGTGGGGAA 2594

Qy       1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
Db       2595 GTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAAGGTGAGGCAGCTGTGCAAACTCCT 2654

Qy       1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db       2655 CCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCTAACCGAGGAGGCCGAGCTCGAACT 2714

Qy       1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db       2715 GGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCACGGCGTGTACTATGACCCCTCCAA 2774

Qy       1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db       2775 GGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGGCCAGTGGACCTATCAGATTTACCA 2834

Qy       1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db       2835 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGGATGAGGGGTGCCCACACTAA 2894

Qy       1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db       2895 CGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGATCACCACCGAAAGCATCGTGATCTG 2954

Qy       1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db       2955 GGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAAGGAAACCTGGGAAACCTGGTGGAC 3014

Qy       1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db       3015 AGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGAGTTCGTCAACACCCCTCCCCTGGT 3074

Qy       1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db       3075 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGTGGGCGCCGAAACCTTCTACGTGGA 3134

Qy       1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db       3135 TGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGCCGGATACGTCACTAACCGGGGCAG 3194

Qy       2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db       3195 ACAGAAGGTTGTCACCCTCACTGACACCACCAACCAGAAGACTGAGCTGCAGGCCATTTA 3254

Qy       2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db       3255 CCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACATCGTGACAGACTCTCAGTATGCCCT 3314

Qy       2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db       3315 GGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTCCGAGCTGGTCAATCAGATCATCGA 3374

Qy       2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db       3375 GCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTGGGTACCCGCCCACAAAGGCATTGG 3434

Qy       2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db       3435 CGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGGCATCAGGAAGGTGCTATTCCTGGA 3494

Qy       2355 CGGCATCGA 2363
Db       3495 TGGCATCGA 3503
```

RESULT 11
US-09-936-572-12
; Sequence 12, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 4642
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP2-codon
; OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence
US-09-936-572-12

Query Match 66.0%; Score 1624.6; DB 4; Length 4642;
Best Local Similarity 81.8%; Pred. No. 1.9e-268;
Matches 1922; Conservative 0; Mismatches 408; Indels 19; Gaps 4;

```
Qy         33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db       1445 GACCAACTCCGCTACCATCATGATGCAGCGCGGCAACTTTCGGAACCAACGCAAGATCGT 1504

Qy         93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db       1505 CAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGCCCGCAACTGCAGGGCCCCTAGGAA 1564
```

```
Qy    153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db   1565 AAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACA 1624

Qy    213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db   1625 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCTTC 1683

Qy    273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGG------CGA 326
Db   1684 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAAGAGACAA 1743

Qy    327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
Db   1744 CAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAGCTTCCCTCA 1803

Qy    381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db   1804 GATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAGGGGGGCAGCTCAAGGAGGC 1863

Qy    441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db   1864 TCTCCTGGACACCGGAGCAGACGACACCGTGCTGGAGGAGATGTCGTTGCCAGGCCGCTG 1923

Qy    501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db   1924 GAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCATCAAGGTGCGCCAGTATGACCAGAT 1983

Qy    561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db   1984 CCTCATCGAAATCTGCGGCCACAAGGCTATCGGTACCGTGCTGGTGGGCCCCACACCCGT 2043

Qy    621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db   2044 CAACATCATCGGACGCAACCTGTTGACGCAGATCGGTTGCACGCTGAACTTCCCCATTAG 2103

Qy    681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db   2104 CCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGGGATGGACGGCCCGAAGGTCAAGCA 2163

Qy    741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db   2164 ATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGTGGAGATTTGCACAGAGATGGAAAA 2223

Qy    801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db   2224 GGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCCGTACAACACGCCGGTGTTCGCAAT 2283

Qy    861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db   2284 CAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 2343

Qy    921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db   2344 CACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCCGCACCCCGCAGGGCTGAAGAAGAA 2403

Qy    981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db   2404 GAAATCCGTGACCGTACTGGATGTGGGTGATGCCTACTTCTCCGTTCCCCTGGACGAAGA 2463

Qy   1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db   2464 CTTCAGGAAGTACACTGCCTTCACAATCCCTTCGATCAACAACGAGACACCGGGGATTCG 2523

Qy   1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db   2524 ATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGGCTCTCCCGCAATCTTCCAGAGTAG 2583

Qy   1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db   2584 CATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAACCCCGACATCGTCATCTATCAGTA 2643

Qy   1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db   2644 CATGGATGACTTGTACGTGGGCTCTGATCTAGAGATAGGGCAGCACCGCACCAAGATCGA 2703

Qy   1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db   2704 GGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGACCACACCCGACAAGAAGCACCAGAA 2763

Qy   1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db   2764 GGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCACCCTGACAAATGGACCGTGCAGCC 2823

Qy   1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db   2824 TATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAACGACATACAGAAGCTGGTGGGGAA 2883

Qy   1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
Db   2884 GTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAAGGTGAGGCAGCTGTGCAAACTCCT 2943

Qy   1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db   2944 CCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCTAACCGAGGAGGCCGAGCTCGAACT 3003

Qy   1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db   3004 GGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCACGGCGTGTACTATGACCCCTCCAA 3063

Qy   1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db   3064 GGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGGCCAGTGGACCTATCAGATTTACCA 3123

Qy   1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db   3124 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGGATGAGGGGTGCCCACACTAA 3183

Qy   1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db   3184 CGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGATCACCACCGAAAGCATCGTGATCTG 3243

Qy   1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db   3244 GGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAAGGAAACCTGGGAAACCTGGTGGAC 3303

Qy   1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db   3304 AGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGAGTTCGTCAACACCCCTCCCCTGGT 3363

Qy   1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db   3364 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGTGGGCGCCGAAACCTTCTACGTGGA 3423

Qy   1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db   3424 TGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGCCGGATACGTCACTAACCGGGGCAG 3483

Qy   2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db   3484 ACAGAAGGTTGTCACCCTCACTGACACCACCAACCAGAAGACTGAGCTGCAGGCCATTTA 3543

Qy   2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db   3544 CCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACATCGTGACAGACTCTCAGTATGCCCT 3603

Qy   2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db   3604 GGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTCCGAGCTGGTCAATCAGATCATCGA 3663

Qy   2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db   3664 GCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTGGGTACCCGCCCACAAAGGCATTGG 3723

Qy   2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
```

```
Db       3724 CGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGGCATCAGGAAGGTGCTATTCCTGGA 3783

Qy       2355 CGGCATCGA 2363

Db       3784 TGGCATCGA 3792


RESULT 12
US-09-552-950-5
; Sequence 5, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
;  APPLICANT: Oxford Biomedica (UK) Limited
;  TITLE OF INVENTION: Anti-Viral Vectors
;  FILE REFERENCE: 674524-2004
;  CURRENT APPLICATION NUMBER: US/09/552,950
;  CURRENT FILING DATE:  2000-04-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 9772
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:pSYNGP
US-09-552-950-5

  Query Match            66.0%;  Score 1624.6;  DB 4;  Length 9772;
  Best Local Similarity  81.8%;  Pred. No. 2.1e-268;
  Matches 1922;  Conservative    0;  Mismatches  408;  Indels   19;  Gaps    4;

Qy         33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGCCCCAAGCGCATCAT 92

Db       2217 GACCAACTCCGCTACCATCATGATGCAGCGCGGCAACTTTCGGAACCAACGCAAGATCGT 2276

Qy         93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152

Db       2277 CAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGCCCGCAACTGCAGGGCCCCTAGGAA 2336

Qy        153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212

Db       2337 AAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACA 2396

Qy        213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272

Db       2397 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCTTC 2455

Qy        273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGG------CGA 326

Db       2456 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAAGAGACAA 2515

Qy        327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380

Db       2516 CAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAGCTTCCCTCA 2575

Qy        381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440

Db       2576 GATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAGGGGGGCAGCTCAAGGAGGC 2635

Qy        441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500

Db       2636 TCTCCTGGACACCGGAGCAGACGACACCGTGCTGGAGGAGATGTCGTTGCCAGGCCGCTG 2695

Qy        501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560

Db       2696 GAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCATCAAGGTGCGCCAGTATGACCAGAT 2755

Qy        561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620

Db       2756 CCTCATCGAAATCTGCGGCCACAAGGCTATCGGTACCGTGCTGGTGGGCCCCACACCCGT 2815

Qy        621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680

Db       2816 CAACATCATCGGACGCAACCTGTTGACGCAGATCGGTTGCACGCTGAACTTCCCCATTAG 2875

Qy        681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740

Db       2876 CCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGGGATGGACGGCCCGAAGGTCAAGCA 2935

Qy        741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800

Db       2936 ATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGTGGAGATTTGCACAGAGATGGAAAA 2995

Qy        801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860

Db       2996 GGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCCGTACAACACGCCGGTGTTCGCAAT 3055

Qy        861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920

Db       3056 CAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 3115

Qy        921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980

Db       3116 CACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCCGCACCCCGCAGGGCTGAAGAAGAA 3175

Qy        981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040

Db       3176 GAAATCCGTGACCGTACTGGATGTGGGTGATGCCTACTTCTCCGTTCCCCTGGACGAAGA 3235

Qy       1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100

Db       3236 CTTCAGGAAGTACACTGCCTTCACAATCCCTTCGATCAACAACGAGACACCGGGGATTCG 3295

Qy       1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160

Db       3296 ATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGGCTCTCCCGCAATCTTCCAGAGTAG 3355

Qy       1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217

Db       3356 CATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAACCCCGACATCGTCATCTATCAGTA 3415

Qy       1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274

Db       3416 CATGGATGACTTGTACGTGGGCTCTGATCTAGAGATAGGGCAGCACCGCACCAAGATCGA 3475

Qy       1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334

Db       3476 GGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGACCACACCCGACAAGAAGCACCAGAA 3535

Qy       1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394

Db       3536 GGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCACCCTGACAAATGGACCGTGCAGCC 3595

Qy       1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454

Db       3596 TATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAACGACATACAGAAGCTGGTGGGGAA 3655

Qy       1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514

Db       3656 GTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAAGGTGAGGCAGCTGTGCAAACTCCT 3715

Qy       1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574

Db       3716 CCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCTAACCGAGGAGGCCGAGCTCGAACT 3775

Qy       1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634

Db       3776 GGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCACGGCGTGTACTATGACCCCTCCAA 3835

Qy       1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694

Db       3836 GGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGGCCAGTGGACCTATCAGATTTACCA 3895

Qy       1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
```

```
Db   3896 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGGATGAGGGGTGCCCACACTAA 3955

Qy   1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db   3956 CGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGATCACCACCGAAAGCATCGTGATCTG 4015

Qy   1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db   4016 GGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAAGGAAACCTGGGAAACCTGGTGGAC 4075

Qy   1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db   4076 AGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGAGTTCGTCAACACCCCTCCCCTGGT 4135

Qy   1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db   4136 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGTGGGCGCCGAAACCTTCTACGTGGA 4195

Qy   1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db   4196 TGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGCCGGATACGTCACTAACCGGGGCAG 4255

Qy   2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db   4256 ACAGAAGGTTGTCACCCTCACTGACACCACCAACCAGAAGACTGAGCTGCAGGCCATTTA 4315

Qy   2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db   4316 CCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACATCGTGACAGACTCTCAGTATGCCCT 4375

Qy   2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db   4376 GGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTCCGAGCTGGTCAATCAGATCATCGA 4435

Qy   2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db   4436 GCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTGGGTACCCGCCCACAAAGGCATTGG 4495

Qy   2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db   4496 CGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGGCATCAGGAAGGTGCTATTCCTGGA 4555

Qy   2355 CGGCATCGA 2363
Db   4556 TGGCATCGA 4564
```

RESULT 13
US-09-872-733A-6
; Sequence 6, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 8366
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: DNA sequence
;   OTHER INFORMATION: of the construct pCMVgagpolBNKan containing a CMV
;   OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin
;   OTHER INFORMATION: resistance gene

US-09-872-733A-6

```
  Query Match             63.6%;  Score 1566;  DB 4;  Length 8366;
  Best Local Similarity   80.2%;  Pred. No. 2e-258;
  Matches 1885;  Conservative    0;  Mismatches  446;  Indels   19;  Gaps    4;

Qy     33 GGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db   1879 GACGAACTCGGCGACCATAATGATGCAGAGAGGCAACTTCCGGAACCAGCGGAAGATCGT 1938

Qy     93 CAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAA 152
Db   1939 CAAGTGCTTCAATTGTGGCAAAGAAGGGCACACCGCCAGGAACTGCCGGGCCCCCCGGAA 1998

Qy    153 GAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCA 212
Db   1999 GAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACA 2058

Qy    213 GGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAG 272
Db   2059 GGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTC 2117

Qy    273 CGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGG------CGA 326
Db   2118 AGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAA 2177

Qy    327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCA 380
Db   2178 CAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCA 2237

Qy    381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db   2238 GATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGATCGGGGGGCAACTCAAGGAAGC 2297

Qy    441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db   2298 GCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAGTTTGCCAGGAAGATG 2357

Qy    501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db   2358 GAAACCAAAAATGATAGGGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGAT 2417

Qy    561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db   2418 ACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGT 2477

Qy    621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db   2478 CAACATAATTGGAAGAAATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAG 2537

Qy    681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db   2538 CCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCA 2597

Qy    741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db   2598 ATGGCCATTGACGAAAGAGAAGATCAAGGCCTTAGTCGAAATCTGTACAGAGATGGAGAA 2657

Qy    801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db   2658 GGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCTACAACACTCCAGTCTTCGCAAT 2717

Qy    861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db   2718 CAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAGCTGAACAAGAG 2777

Qy    921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db   2778 AACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCACATCCCGCTGGGTTGAAGAAGAA 2837

Qy    981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db   2838 GAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACTTCTCCGTTCCCTTGGACGAGGA 2897
```

```
Qy       1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCGGCATCCG 1100
Db       2898 CTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCAACAACGAGACACCAGGCATCCG 2957

Qy       1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db       2958 CTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGATCACCAGCCATCTTTCAAAGCAG 3017

Qy       1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db       3018 CATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACCCAGACATCGTGATCTATCAGTA 3077

Qy       1218 ---GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db       3078 CATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCGGGCAGCACAGGACCAAGATCGA 3137

Qy       1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db       3138 GGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCACACCAGACAAGAAGCACCAGAA 3197

Qy       1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db       3198 GGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCATCCTGACAAGTGGACAGTGCAGCC 3257

Qy       1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db       3258 CATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACGACATACAGAAGCTCGTGGGCAA 3317

Qy       1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
Db       3318 GTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAGTTAGGCAGCTGTGCAAGCTGCT 3377

Qy       1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db       3378 TCGAGGAACCAAGGCACTGACAGAAGTGATCCCACTGACAGAGGAAGCAGAGCTAGAACT 3437

Qy       1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db       3438 GGCAGAGAACCGAGAGATCCTGAAGGAGCCAGTACATGGAGTGTACTACGACCCAAGCAA 3497

Qy       1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db       3498 GGACCTGATCGCAGAGATCCAGAAGCAGGGGCAAGGCCAATGGACCTACCAAATCTACCA 3557

Qy       1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
Db       3558 GGAGCCCTTCAAGAACCTGAAGACAGGCAAGTACGCAAGGATGAGGGGTGCCCACACCAA 3617

Qy       1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db       3618 CGATGTGAAGCAGCTGACAGAGGCAGTGCAGAAGATCACCACAGAGAGCATCGTGATCTG 3677

Qy       1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db       3678 GGGCAAGACTCCCAAGTTCAAGCTGCCCATACAGAAGGAGACATGGGAGACATGGTGGAC 3737

Qy       1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db       3738 CGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTCGTGAACACCCCTCCCTTGGT 3797

Qy       1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db       3798 GAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGGGAGCAGAGACCTTCTACGTGGA 3857

Qy       1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db       3858 TGGGGCAGCCAACAGGGAGACCAAGCTGGGCAAGGCAGGCTACGTGACCAACCGAGGACG 3917

Qy       2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db       3918 ACAGAAAGTGGTGACCCTGACTGACACCACCAACCAGAAGACTGAGCTGCAAGCCATCTA 3977

Qy       2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db       3978 CCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCGTGACAGACTCACAGTACGCACT 4037

Qy       2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db       4038 GGGCATCATCCAAGCACAACCAGACCAATCCGAGTCAGAGCTGGTGAACCAGATCATCGA 4097

Qy       2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db       4098 GCAGCTGATCAAGAAGGAGAAAGTGTACCTGGCATGGGTACCAGCACACAAAGGAATTGG 4157

Qy       2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGA 2354
Db       4158 AGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGGATCCGGAAGGTGCTGTTCCTGGA 4217

Qy       2355 CGGCATCGAT 2364
Db       4218 CGGGATCGAT 4227
```

```
RESULT 14
US-09-872-733A-1
; Sequence 1, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
;  APPLICANT: The Government of the United States of America, as
;  TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
;  TITLE OF INVENTION:  SIV ENV GENES
;  FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
;  CURRENT APPLICATION NUMBER: US/09/872,733A
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: PCT/US00/34985
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/173,036
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4338
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Mutated Human
;   OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene
US-09-872-733A-1

  Query Match             62.1%;  Score 1530.6;  DB 4;  Length 4338;
  Best Local Similarity   80.1%;  Pred. No. 2e-252;
  Matches 1905;  Conservative    0;  Mismatches  444;  Indels   28;  Gaps    8;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70
Db       1085 TGGCCGAGGCGATGAGCCAGGTGACGAACTCGGCGACCATAATGATGCAGAGAGGCAACT 1144

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db       1145 TCCGGAACCAGCGGAAGATCGTCAAGTGCTTCAATTGTGGCAAAGAAGGGCACACCGCCA 1204

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       1205 GGAACTGCCGGGCCCCCCGGAAGAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAA 1264

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       1265 TGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAG 1323

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       1324 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 1383

Qy        311 TGCAGGTGCGCGG-----CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
```

```
Db   1384 TTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAA 1443

Qy    365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db   1444 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 1503

Qy    419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db   1504 TCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAG 1563

Qy    479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db   1564 AAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGGATCGGGGGCTTCATCA 1623

Qy    539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db   1624 AGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 1683

Qy    599 TGCTGATCGGCCCCACC----CCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654
Db   1684 TATTAGTAGGACCTACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATC 1743

Qy    655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714
Db   1744 GGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCG 1803

Qy    715 GGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTG 774
Db   1804 GGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCTTA 1863

Qy    775 ACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAAC 834
Db   1864 GTCGAAATCTGTACAGAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAAC 1923

Qy    835 CCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTG 894
Db   1924 CCCTACAACACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTG 1983

Qy    895 GTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954
Db   1984 GTGGACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATC 2043

Qy    955 CCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCC 1014
Db   2044 CCACATCCCGCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCC 2103

Qy   1015 TACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGC 1074
Db   2104 TACTTCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGC 2163

Qy   1075 ATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAG 1134
Db   2164 ATCAACAACGAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAG 2223

Qy   1135 GGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGC 1194
Db   2224 GGATCACCAGCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAA 2283

Qy   1195 AACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAG 1248
Db   2284 AACCCAGACATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAG 2343

Qy   1249 ATC-GGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTT 1307
Db   2344 ATCGGGGCAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACT 2403

Qy   1308 CACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCT 1367
Db   2404 GACCACACCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACT 2463

Qy   1368 GCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGT 1427
Db   2464 GCATCCTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGT 2523

Qy   1428 GAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCAT 1487
Db   2524 GAACGACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCAT 2583

Qy   1488 CAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCC 1547
Db   2584 CAAAGTTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCC 2643

Qy   1548 CCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGT 1607
Db   2644 ACTGACAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGT 2703

Qy   1608 GCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCA 1667
Db   2704 ACATGGAGTGTACTACGACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCA 2763

Qy   1668 CGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTA 1727
Db   2764 AGGCCAATGGACCTACCAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTA 2823

Qy   1728 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1787
Db   2824 CGCAAGGATGAGGGGTGCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAA 2883

Qy   1788 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1847
Db   2884 GATCACCACAGAGAGCATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACA 2943

Qy   1848 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1907
Db   2944 GAAGGAGACATGGGAGACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTG 3003

Qy   1908 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1967
Db   3004 GGAGTTCGTGAACACCCCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCAT 3063

Qy   1968 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2027
Db   3064 CGTGGGAGCAGAGACCTTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAA 3123

Qy   2028 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2087
Db   3124 GGCAGGCTACGTGACCAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAA 3183

Qy   2088 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2147
Db   3184 CCAGAAGACTGAGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAA 3243

Qy   2148 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2207
Db   3244 CATCGTGACAGACTCACAGTACG-CATGGGCATCATCCAAGCACAACCAGACCAATCCGA 3302

Qy   2208 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2267
Db   3303 GTCAGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGC 3362

Qy   2268 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2327
Db   3363 ATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGC 3422

Qy   2328 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db   3423 TGGGATCCGGAAGGTGCTGTTCCTGGACGGGATCGAT 3459
```

RESULT 15
US-09-952-060-1
; Sequence 1, Application US/09952060
; Patent No. 6733993
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.
; APPLICANT: Youil, Rima

```
; APPLICANT:  Bett, Andrew J.
; APPLICANT:  Chen, Ling
; APPLICANT:  Kaslow, David C.
; APPLICANT:  Shiver, John W.
; APPLICANT:  Toner, Timothy J.
; APPLICANT:  Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION:  VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION:  MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE:  2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2577
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Codon optimized DNA encoding modified wt pol
US-09-952-060-1

  Query Match             53.2%;  Score 1309.6;  DB 4;  Length 2577;
  Best Local Similarity   86.2%;  Pred. No. 9.6e-215;
  Matches 1463;  Conservative    0;  Mismatches  229;  Indels    6;  Gaps    1;

Qy        672 CCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAA 731
Db         15 CCCCATCTCCCCCATTGAGACTGTGCCTGTGAAGCTGAAGCCTGGCATGGATGGCCCCAA 74

Qy        732 GGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGA 791
Db         75 GGTGAAGCAGTGGCCCCTGACTGAGGAGAAGATCAAGGCCCTGGTGGAAATCTGCACTGA 134

Qy        792 GATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGT 851
Db        135 GATGGAGAAGGAGGGCAAAATCTCCAAGATTGGCCCCGAGAACCCCTACAACACCCCTGT 194

Qy        852 GTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCT 911
Db        195 GTTTGCCATCAAGAAGAAGGACTCCACCAAGTGGAGGAAGCTGGTGGACTTCAGGGAGCT 254

Qy        912 GAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCT 971
Db        255 GAACAAGAGGACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCTGGCCT 314

Qy        972 GAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCT 1031
Db        315 GAAGAAGAAGAAGTCTGTGACTGTGCTGGATGTGGGGGATGCCTACTTCTCTGTGCCCCT 374

Qy       1032 GGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCC 1091
Db        375 GGATGAGGACTTCAGGAAGTACACTGCCTTCACCATCCCCTCCATCAACAATGAGACCCC 434

Qy       1092 CGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTT 1151
Db        435 TGGCATCAGGTACCAGTACAATGTGCTGCCCCAGGGCTGGAAGGGCTCCCCTGCCATCTT 494

Qy       1152 CCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGAT 1211
Db        495 CCAGTCCTCCATGACCAAGATCCTGGAGCCCTTCAGGAAGCAGAACCCTGACATTGTGAT 554

Qy       1212 CTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGC 1265
Db        555 CTACCAGTACATGGATGACCTGTATGTGGGCTCTGACCTGGAGATTGGGCAGCACAGGAC 614

Qy       1266 CAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAA 1325
Db        615 CAAGATTGAGGAGCTGAGGCAGCACCTGCTGAGGTGGGGCCTGACCACCCCTGACAAGAA 674

Qy       1326 GCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGAC 1385
Db        675 GCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTATGAGCTGCACCCCGACAAGTGGAC 734

Qy       1386 CGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCT 1445
Db        735 TGTGCAGCCCATTGTGCTGCCTGAGAAGGACTCCTGGACTGTGAATGACATCCAGAAGCT 794

Qy       1446 GGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTG 1505
Db        795 GGTGGGCAAGCTGAACTGGGCCTCCCAAATCTACCCTGGCATCAAGGTGAGGCAGCTGTG 854

Qy       1506 CAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGA 1565
Db        855 CAAGCTGCTGAGGGGCACCAAGGCCCTGACTGAGGTGATCCCCCTGACTGAGGAGGCTGA 914

Qy       1566 GCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGA 1625
Db        915 GCTGGAGCTGGCTGAGAACAGGGAGATCCTGAAGGAGCCTGTGCATGGGGTGTACTATGA 974

Qy       1626 CCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCA 1685
Db        975 CCCCTCCAAGGACCTGATTGCTGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCA 1034

Qy       1686 GATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGC 1745
Db       1035 AATCTACCAGGAGCCCTTCAAGAACCTGAAGACTGGCAAGTATGCCAGGATGAGGGGGGC 1094

Qy       1746 CCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCAT 1805
Db       1095 CCACACCAATGATGTGAAGCAGCTGACTGAGGCTGTGCAGAAGATCACCACTGAGTCCAT 1154

Qy       1806 CGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGAC 1865
Db       1155 TGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGACCTGGGAGAC 1214

Qy       1866 CTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCC 1925
Db       1215 CTGGTGGACTGAGTACTGGCAGGCCACCTGGATCCCTGAGTGGGAGTTTGTGAACACCCC 1274

Qy       1926 CCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTT 1985
Db       1275 CCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATTGTGGGGGCTGAGACCTT 1334

Qy       1986 CTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGA 2045
Db       1335 CTATGTGGATGGGGCTGCCAACAGGGAGACCAAGCTGGGCAAGGCTGGCTATGTGACCAA 1394

Qy       2046 CCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCA 2105
Db       1395 CAGGGGCAGGCAGAAGGTGGTGACCCTGACTGACACCACCAACCAGAAGACTGAGCTCCA 1454

Qy       2106 GGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCA 2165
Db       1455 GGCCATCTACCTGGCCCTCCAGGACTCTGGCCTGGAGGTGAACATTGTGACTGACTCCCA 1514

Qy       2166 GTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCA 2225
Db       1515 GTATGCCCTGGGCATCATCCAGGCCCAGCCTGATCAGTCTGAGTCTGAGCTGGTGAACCA 1574

Qy       2226 GATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAA 2285
Db       1575 GATCATTGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGCCTGCCCACAA 1634

Qy       2286 GGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCT 2345
Db       1635 GGGCATTGGGGGCAATGAGCAGGTGGACAAGCTGGTGTCTGCTGGCATCAGGAAGGTGCT 1694
```

```
Qy        2346 GTTCCTGGACGGCATCGA 2363
               ||||||||| ||||| ||
Db        1695 GTTCCTGGATGGCATTGA 1712
```

Search completed: June 2, 2005, 03:13:05
Job time : 401 secs

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 09:47:50 ; Search time 388.052 Seconds
(without alignments)
10360.282 Million cell updates/sec

Title: US-09-610-313B-32
Perfect score: 2457
Sequence: 1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2040	83.0	2300	4	US-09-475-515-83	Sequence 83, Appl
2	2019.2	82.2	2306	4	US-09-475-515-82	Sequence 82, Appl
3	1998.4	81.3	2312	4	US-09-475-515-84	Sequence 84, Appl
4	1915.8	78.0	4319	4	US-09-475-515-6	Sequence 6, Appli
5	1872.8	76.2	2299	4	US-09-475-515-81	Sequence 81, Appl
6	1852	75.4	2305	4	US-09-475-515-80	Sequence 80, Appl
7	1626.6	66.2	4307	4	US-09-552-950-2	Sequence 2, Appli
8	1607.4	65.4	4307	4	US-09-936-572-2	Sequence 2, Appli
9	1599.4	65.1	4327	4	US-09-936-572-14	Sequence 14, Appl
10	1599.4	65.1	4353	4	US-09-936-572-13	Sequence 13, Appl
11	1599.4	65.1	4642	4	US-09-936-572-12	Sequence 12, Appl
12	1599.4	65.1	9772	4	US-09-552-950-5	Sequence 5, Appli
13	1539.2	62.6	8366	4	US-09-872-733A-6	Sequence 6, Appli
14	1503.8	61.2	4338	4	US-09-872-733A-1	Sequence 1, Appli
15	1282.8	52.2	2577	4	US-09-952-060-1	Sequence 1, Appli
16	1282.8	52.2	2650	4	US-09-952-060-5	Sequence 5, Appli
17	1280.2	52.1	4053	4	US-09-952-060-34	Sequence 34, Appl
18	1278	52.0	2577	4	US-09-952-060-3	Sequence 3, Appli
19	1278	52.0	2650	4	US-09-952-060-7	Sequence 7, Appli
20	1278	52.0	38519	4	US-09-952-060-28	Sequence 28, Appl
21	1178.4	48.0	9010	4	US-09-184-418C-8	Sequence 8, Appli
22	1150.6	46.8	9913	4	US-09-827-688-11	Sequence 11, Appl
23	1147.4	46.7	8972	4	US-09-184-418C-9	Sequence 9, Appli
24	1140.2	46.4	8959	4	US-09-184-418C-11	Sequence 11, Appl
25	1137	46.3	2467	4	US-09-872-733A-3	Sequence 3, Appli
26	1116.8	45.5	8992	4	US-09-184-418C-4	Sequence 4, Appli
27	1081.6	44.0	2601	3	US-09-117-217-7	Sequence 7, Appli
28	1081.6	44.0	2601	3	US-09-117-217-9	Sequence 9, Appli
29	1081.6	44.0	2601	3	US-09-117-217-11	Sequence 11, Appl
30	1081.6	44.0	2601	3	US-09-117-217-13	Sequence 13, Appl
31	1081.6	44.0	2601	4	US-09-735-487-7	Sequence 7, Appli
32	1081.6	44.0	2601	4	US-09-735-487-9	Sequence 9, Appli
33	1081.6	44.0	2601	4	US-09-735-487-11	Sequence 11, Appl
34	1081.6	44.0	2601	4	US-09-735-487-13	Sequence 13, Appl
35	1081.6	44.0	4307	4	US-09-552-950-1	Sequence 1, Appli
36	1081.6	44.0	4307	4	US-09-936-572-1	Sequence 1, Appli
37	1081.6	44.0	9719	4	US-09-700-304-1	Sequence 1, Appli
38	1080	44.0	9050	4	US-09-184-418C-7	Sequence 7, Appli
39	1075.2	43.8	7399	2	US-08-418-848A-9	Sequence 9, Appli
40	1075.2	43.8	9709	2	US-08-188-583-5	Sequence 5, Appli
41	1075.2	43.8	9709	3	US-08-388-353-1	Sequence 1, Appli
42	1075.2	43.8	9709	3	US-08-488-551B-1	Sequence 1, Appli
43	1075.2	43.8	9709	3	US-09-309-572-15	Sequence 15, Appl
44	1075.2	43.8	9709	4	US-09-718-096-15	Sequence 15, Appl
45	1075.2	43.8	12479	4	US-09-318-138-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-475-515-83
; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt.YMWM
US-09-475-515-83

```
  Query Match             83.0%;  Score 2040;  DB 4;  Length 2300;
  Best Local Similarity   93.6%;  Pred. No. 0;
  Matches 2153;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              |||||||||||||||  |||||||||||||||||||||  ||||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              ||||||||||| ||||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| ||||||||||||| ||      ||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240
```

```
Qy    398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db    241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy    458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db    301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy    518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db    361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy    578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db    421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy    638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db    481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy    698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db    541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy    758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db    601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy    818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db    661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy    878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db    721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy    938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db    781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy    998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db    841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy   1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db    901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy   1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db    961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy   1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db   1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy   1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db   1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy   1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1357
Db   1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1200

Qy   1358 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGA 1417
Db   1201 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA 1260

Qy   1418 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477
Db   1261 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG 1320

Qy   1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG 1537
Db   1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGA 1380

Qy   1538 TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGC 1597
Db   1381 TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC 1440

Qy   1598 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1657
Db   1441 CCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1500

Qy   1658 GCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1717
Db   1501 GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1560

Qy   1718 AGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1777
Db   1561 AGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1620

Qy   1778 AGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCA 1837
Db   1621 AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA 1680

Qy   1838 TCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCG 1897
Db   1681 TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG 1740

Qy   1898 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1957
Db   1741 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1800

Qy   1958 CCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCG 2017
Db   1801 CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG 1860

Qy   2018 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCA 2077
Db   1861 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA 1920

Qy   2078 CCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG 2137
Db   1921 CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG 1980

Qy   2138 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2197
Db   1981 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2040

Qy   2198 GCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2257
Db   2041 GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2100

Qy   2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
Db   2101 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2160

Qy   2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377
Db   2161 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2220

Qy   2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437
Db   2221 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2280

Qy   2438 GGGCTAGCACCGGTGAATTC 2457
Db   2281 GGGCTAGCACCGGTGAATTC 2300
```

RESULT 2
US-09-475-515-82
; Sequence 82, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:

```
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
;   LENGTH: 2306
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(-).protmod.RTopt.YM
US-09-475-515-82

  Query Match             82.2%;  Score 2019.2;  DB 4;  Length 2306;
  Best Local Similarity   93.2%;  Pred. No. 0;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1355
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy       1356 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1411
Db       1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy       1412 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1471
Db       1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy       1472 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1531
Db       1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy       1532 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1591
Db       1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy       1592 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1651
Db       1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy       1652 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1711
Db       1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy       1712 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1771
Db       1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy       1772 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1831
Db       1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy       1832 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1891
Db       1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy       1892 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1951
```

```
Db       1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy       1952 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2011

Db       1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy       2012 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2071

Db       1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy       2072 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2131

Db       1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy       2132 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2191

Db       1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy       2192 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2251

Db       2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy       2252 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2311

Db       2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy       2312 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2371

Db       2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy       2372 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2431

Db       2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy       2432 TTCCCGGGGCTAGCACCGGTGAATTC 2457

Db       2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

RESULT 3
US-09-475-515-84
; Sequence 84, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 84
;   LENGTH: 2312
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(-).protmod.RTopt(+)
US-09-475-515-84

```
  Query Match             81.3%;  Score 1998.4;  DB 4;  Length 2312;
  Best Local Similarity   92.9%;  Pred. No. 0;
  Matches 2147;  Conservative    0;  Mismatches  141;  Indels   24;  Gaps    4;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757

Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080
```

```
Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGC 1351
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1405
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1406 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1465
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1466 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1525
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1526 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1585
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1586 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1645
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1646 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1705
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1706 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1765
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1766 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1825
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1826 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1885
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1886 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1945
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1946 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2005
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2006 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2065
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2066 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2125
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2126 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2185
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2186 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2245
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2246 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2305
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2306 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2365
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2366 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2425
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2426 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

RESULT 4
US-09-475-515-6
; Sequence 6, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: HIV-Gag-polymerase
US-09-475-515-6

Query Match 78.0%; Score 1915.8; DB 4; Length 4319;
Best Local Similarity 89.4%; Pred. No. 0;
Matches 2140; Conservative 0; Mismatches 227; Indels 28; Gaps 6;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70
Db       1100 TGGCCGAGGCGATGAGCCAGGTGACGAACCCGGCGACCATCATGATGCAGCGCGGCAACT 1159

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db       1160 TCCGCAACCAGCGGAAGACCGTCAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCA 1219

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       1220 GGAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGCGCTGCGGCCGCGAAGGACACCAAA 1279

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       1280 TGAAAGATTGCACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAG 1338

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       1339 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 1398

Qy        311 TGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db       1399 TTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 1458

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db       1459 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 1518
```

```
Qy    419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db   1519 TCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCGCCGACGACACCGTGCTGGAGG 1578

Qy    479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db   1579 AGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGGATCGGGGGCTTCATCA 1638

Qy    539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db   1639 AGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCGGCCACAAGGCCATCGGCACCG 1698

Qy    599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
Db   1699 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 1758

Qy    659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
Db   1759 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGCCCGTGAAGCTGAAGCCGGGGA 1818

Qy    719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
Db   1819 TGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 1878

Qy    779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
Db   1879 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 1938

Qy    839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
Db   1939 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 1998

Qy    899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
Db   1999 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 2058

Qy    959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
Db   2059 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 2118

Qy   1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
Db   2119 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 2178

Qy   1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
Db   2179 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 2238

Qy   1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db   2239 GCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 2298

Qy   1199 CCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCG 1252
Db   2299 CCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 2358

Qy   1253 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1312
Db   2359 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 2418

Qy   1313 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACC 1366
Db   2419 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGCTACGAGCTGCACC 2478

Qy   1367 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1426
Db   2479 CCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 2538

Qy   1427 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1486
Db   2539 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 2598

Qy   1487 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1546
Db   2599 TGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGATCCCCCTGA 2658

Qy   1547 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1606
Db   2659 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 2718

Qy   1607 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1666
Db   2719 AGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCAGGGCC 2778

Qy   1667 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1726
Db   2779 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 2838

Qy   1727 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1786
Db   2839 GCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGGTGA 2898

Qy   1787 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1846
Db   2899 GCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCATCCAGAAGG 2958

Qy   1847 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1906
Db   2959 AGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 3018

Qy   1907 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1966
Db   3019 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCGTGG 3078

Qy   1967 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2026
Db   3079 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 3138

Qy   2027 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2086
Db   3139 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGAGCATCGCCGACACCACCAACCAGA 3198

Qy   2087 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2146
Db   3199 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 3258

Qy   2147 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2206
Db   3259 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 3318

Qy   2207 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2266
Db   3319 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 3378

Qy   2267 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2326
Db   3379 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGAGCGCCGGCA 3438

Qy   2327 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
Db   3439 TCCGCAAGGTGCTGTTCCTGAACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 3493
```

RESULT 5
US-09-475-515-81
; Sequence 81, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine

```
;  APPLICANT: SELBY, Mark
;  APPLICANT: WALKER, Christopher
;  TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
;  TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
;  FILE REFERENCE: 1621.002
;  CURRENT APPLICATION NUMBER: US/09/475,515A
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
;   LENGTH: 2299
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(+).proinact.RTopt.YMWM
US-09-475-515-81

  Query Match             76.2%;  Score 1872.8;  DB 4;  Length 2299;
  Best Local Similarity   89.3%;  Pred. No. 2.2e-310;
  Matches 2055;  Conservative    0;  Mismatches  232;  Indels   13;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTT-AGGG 59

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         60 AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA 119

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        120 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG 179

Qy        344 CCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        180 CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        240 GACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAG 299

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG 359

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        360 GGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTG 419

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        480 ATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGC 539

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        540 CCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        600 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 659

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        660 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 719

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        720 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 779

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        780 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 839

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        840 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 899

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        900 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 959

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        960 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1019

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1020 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1079

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1080 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1139

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1357
Db       1140 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1199

Qy       1358 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGA 1417
Db       1200 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA 1259

Qy       1418 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477
Db       1260 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG 1319

Qy       1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG 1537
Db       1320 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGA 1379

Qy       1538 TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGC 1597
Db       1380 TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC 1439

Qy       1598 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1657
Db       1440 CCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1499

Qy       1658 GCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1717
Db       1500 GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1559

Qy       1718 AGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1777
Db       1560 AGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1619

Qy       1778 AGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCA 1837
Db       1620 AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA 1679

Qy       1838 TCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCG 1897
Db       1680 TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG 1739

Qy       1898 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1957
Db       1740 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1799

Qy       1958 CCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCG 2017
Db       1800 CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG 1859
```

```
Qy       2018 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCA 2077
Db       1860 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA 1919

Qy       2078 CCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG 2137
Db       1920 CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG 1979

Qy       2138 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2197
Db       1980 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2039

Qy       2198 GCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2257
Db       2040 GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2099

Qy       2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
Db       2100 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2159

Qy       2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377
Db       2160 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2219

Qy       2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437
Db       2220 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2279

Qy       2438 GGGCTAGCACCGGTGAATTC 2457
Db       2280 GGGCTAGCACCGGTGAATTC 2299
```

RESULT 6
US-09-475-515-80
; Sequence 80, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
;   LENGTH: 2305
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: FS(+).proinact.RTopt.YM
US-09-475-515-80

```
  Query Match             75.4%;  Score 1852;  DB 4;  Length 2305;
  Best Local Similarity   89.0%;  Pred. No. 7.5e-307;
  Matches 2052;  Conservative    0;  Mismatches  235;  Indels   19;  Gaps    4;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTT-AGGG 59

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         60 AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA 119

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        120 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG 179

Qy        344 CCGGCGCCGAGCGCCAGGGCA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        180 CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        240 GACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAG 299

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        300 CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG 359

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        360 GGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTG 419

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        480 ATCTGTTGACCCAGATCGGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGC 539

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        540 CCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        600 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 659

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        660 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 719

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        720 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 779

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        780 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 839

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        840 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 899

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        900 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 959

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        960 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1019

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1020 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1079

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1080 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1139

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1355
Db       1140 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1199
```

```
Qy   1356 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1411
Db   1200 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1259

Qy   1412 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1471
Db   1260 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1319

Qy   1472 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1531
Db   1320 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1379

Qy   1532 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1591
Db   1380 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1439

Qy   1592 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1651
Db   1440 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1499

Qy   1652 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1711
Db   1500 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1559

Qy   1712 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1771
Db   1560 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1619

Qy   1772 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1831
Db   1620 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1679

Qy   1832 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1891
Db   1680 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1739

Qy   1892 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1951
Db   1740 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1799

Qy   1952 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2011
Db   1800 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1859

Qy   2012 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2071
Db   1860 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1919

Qy   2072 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2131
Db   1920 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1979

Qy   2132 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2191
Db   1980 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2039

Qy   2192 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2251
Db   2040 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2099

Qy   2252 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2311
Db   2100 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2159

Qy   2312 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2371
Db   2160 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2219

Qy   2372 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2431
Db   2220 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2279

Qy   2432 TTCCCGGGGCTAGCACCGGTGAATTC 2457
Db   2280 TTCCCGGGGCTAGCACCGGTGAATTC 2305
```

```
RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
;  APPLICANT: Oxford Biomedica (UK) Limited
;  TITLE OF INVENTION: Anti-Viral Vectors
;  FILE REFERENCE: 674524-2004
;  CURRENT APPLICATION NUMBER: US/09/552,950
;  CURRENT FILING DATE:  2000-04-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4307
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:gagpol-SYNgp - codon
;   OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

  Query Match             66.2%;  Score 1626.6;  DB 4;  Length 4307;
  Best Local Similarity   82.2%;  Pred. No. 2.1e-268;
  Matches 1951;  Conservative    0;  Mismatches  394;  Indels   28;  Gaps    6;

Qy     12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db   1086 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1145

Qy     69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db   1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1205

Qy    129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db   1206 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGCTGGAAATGCGGCAAGGAAGGCCACCA 1265

Qy    189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db   1266 GATGAAAGACTGTACGGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACA 1324

Qy    249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db   1325 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCCGCGCCAACAGCCCCACCCGCCGCGA 1384

Qy    309 GCTGCAGGT------GCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db   1385 GCTTCAGGTCTGGGGTCGCGACAACAACTCCCCCTCCGAAGCAGGAGCCGACCGCCAGGG 1444

Qy    363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db   1445 CACGGTGTCCTTCAACTTCCCTCAGGTCACGCTTTGGCAGCGACCCCTCGTCACCATCAA 1504

Qy    417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db   1505 GATCGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1564

Qy    477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db   1565 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1624

Qy    537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db   1625 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1684

Qy    597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db   1685 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 1744
```

```
Qy     657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db    1745 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 1804

Qy     717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db    1805 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 1864

Qy     777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db    1865 GGAGATTTGCACAGAGATGGAAAAGGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCC 1924

Qy     837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db    1925 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 1984

Qy     897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db    1985 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2044

Qy     957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db    2045 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2104

Qy    1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db    2105 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2164

Qy    1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db    2165 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2224

Qy    1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db    2225 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2284

Qy    1197 CCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250
Db    2285 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2344

Qy    1251 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1310
Db    2345 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2404

Qy    1311 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCA 1364
Db    2405 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 2464

Qy    1365 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1424
Db    2465 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2524

Qy    1425 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1484
Db    2525 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2584

Qy    1485 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1544
Db    2585 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2644

Qy    1545 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1604
Db    2645 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 2704

Qy    1605 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1664
Db    2705 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 2764

Qy    1665 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1724
Db    2765 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 2824

Qy    1725 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1784
Db    2825 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 2884

Qy    1785 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1844
Db    2885 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 2944

Qy    1845 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1904
Db    2945 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3004

Qy    1905 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1964
Db    3005 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3064

Qy    1965 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2024
Db    3065 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3124

Qy    2025 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2084
Db    3125 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3184

Qy    2085 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2144
Db    3185 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3244

Qy    2145 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2204
Db    3245 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3304

Qy    2205 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2264
Db    3305 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3364

Qy    2265 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2324
Db    3365 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3424

Qy    2325 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2357
Db    3425 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3457
```

```
RESULT 8
US-09-936-572-2
; Sequence 2, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
;  APPLICANT: UDEN, MARK
;  APPLICANT:  MITROPHANOUS, KYRIACOS
;  TITLE OF INVENTION: ANTI-VIRAL VECTORS
;  FILE REFERENCE: 078883/0137
;  CURRENT APPLICATION NUMBER: US/09/936,572
;  CURRENT FILING DATE:  2001-12-11
;  PRIOR APPLICATION NUMBER: PCT/GB00/01002
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: GB 9906177.2
;  PRIOR FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4307
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: gagpol-SYNgp-codon optimised gagpol sequence
US-09-936-572-2

  Query Match             65.4%;  Score 1607.4;  DB 4;  Length 4307;
  Best Local Similarity   81.7%;  Pred. No. 3.9e-265;
```

```
Matches 1939; Conservative    0; Mismatches  406; Indels   28; Gaps    6;

Qy        12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db      1086 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1145

Qy        69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db      1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1205

Qy       129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db      1206 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGCTGGAAATGCGGCAAGGAAGGCCACCA 1265

Qy       189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db      1266 GATGAAAGACTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACA 1324

Qy       249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db      1325 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGA 1384

Qy       309 GCTGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db      1385 GCTTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGG 1444

Qy       363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db      1445 AACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATAAA 1504

Qy       417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db      1505 GATAGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1564

Qy       477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db      1565 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1624

Qy       537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db      1625 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1684

Qy       597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db      1685 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 1744

Qy       657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db      1745 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 1804

Qy       717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db      1805 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 1864

Qy       777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db      1865 GGAGATTTGCACAGAGATGGAAAAGGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCC 1924

Qy       837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db      1925 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 1984

Qy       897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db      1985 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2044

Qy       957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db      2045 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2104

Qy      1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db      2105 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2164

Qy      1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db      2165 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2224

Qy      1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db      2225 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2284

Qy      1197 CCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250
Db      2285 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2344

Qy      1251 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1310
Db      2345 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2404

Qy      1311 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCA 1364
Db      2405 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 2464

Qy      1365 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1424
Db      2465 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2524

Qy      1425 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1484
Db      2525 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2584

Qy      1485 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1544
Db      2585 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2644

Qy      1545 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1604
Db      2645 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 2704

Qy      1605 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1664
Db      2705 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 2764

Qy      1665 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1724
Db      2765 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 2824

Qy      1725 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1784
Db      2825 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 2884

Qy      1785 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1844
Db      2885 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 2944

Qy      1845 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1904
Db      2945 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3004

Qy      1905 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1964
Db      3005 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3064

Qy      1965 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2024
Db      3065 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3124

Qy      2025 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2084
Db      3125 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3184

Qy      2085 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2144
Db      3185 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3244
```

```
Qy       2145 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2204
Db       3245 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3304

Qy       2205 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2264
Db       3305 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3364

Qy       2265 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2324
Db       3365 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3424

Qy       2325 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2357
Db       3425 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3457
```

RESULT 9
US-09-936-572-14
; Sequence 14, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 4327
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP4-codon
; OTHER INFORMATION: optimised HIV-1 gagpol with 20bp of the leader
; OTHER INFORMATION: sequence of HIV-1
US-09-936-572-14

```
  Query Match             65.1%;  Score 1599.4;  DB 4;  Length 4327;
  Best Local Similarity   81.5%;  Pred. No. 9e-264;
  Matches 1934;  Conservative    0;  Mismatches  411;  Indels   28;  Gaps    6;

Qy         12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db       1106 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1165

Qy         69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db       1166 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1225

Qy        129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db       1226 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCA 1285

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db       1286 AATGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACA 1344

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db       1345 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGA 1404

Qy        309 GCTGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db       1405 GCTTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGG 1464

Qy        363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db       1465 AACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAA 1524

Qy        417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db       1525 GATAGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1584

Qy        477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db       1585 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1644

Qy        537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db       1645 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1704

Qy        597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db       1705 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 1764

Qy        657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db       1765 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 1824

Qy        717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db       1825 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 1884

Qy        777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db       1885 GGAGATTTGCACAGAGATGGAAAAGGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCC 1944

Qy        837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db       1945 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 2004

Qy        897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db       2005 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2064

Qy        957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db       2065 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2124

Qy       1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db       2125 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2184

Qy       1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db       2185 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2244

Qy       1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db       2245 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2304

Qy       1197 CCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250
Db       2305 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2364

Qy       1251 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1310
Db       2365 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2424

Qy       1311 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCA 1364
Db       2425 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGTTACGAGCTGCA 2484

Qy       1365 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1424
Db       2485 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2544

Qy       1425 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1484
```

```
Db       2545 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2604

Qy       1485 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1544
Db       2605 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2664

Qy       1545 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1604
Db       2665 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 2724

Qy       1605 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1664
Db       2725 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 2784

Qy       1665 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1724
Db       2785 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 2844

Qy       1725 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1784
Db       2845 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 2904

Qy       1785 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1844
Db       2905 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 2964

Qy       1845 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1904
Db       2965 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3024

Qy       1905 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1964
Db       3025 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3084

Qy       1965 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2024
Db       3085 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3144

Qy       2025 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2084
Db       3145 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3204

Qy       2085 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2144
Db       3205 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3264

Qy       2145 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2204
Db       3265 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3324

Qy       2205 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2264
Db       3325 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3384

Qy       2265 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2324
Db       3385 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3444

Qy       2325 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2357
Db       3445 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3477
```

RESULT 10
US-09-936-572-13
; Sequence 13, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 4353
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP3-codon
; OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence from
; OTHER INFORMATION: the major splice donor
US-09-936-572-13

```
Query Match            65.1%;  Score 1599.4;  DB 4;  Length 4353;
Best Local Similarity  81.5%;  Pred. No. 9e-264;
Matches 1934;  Conservative    0;  Mismatches  411;  Indels   28;  Gaps    6;

Qy         12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db       1132 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1191

Qy         69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db       1192 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1251

Qy        129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db       1252 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCA 1311

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db       1312 AATGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACA 1370

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db       1371 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGA 1430

Qy        309 GCTGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db       1431 GCTTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGG 1490

Qy        363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db       1491 AACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAA 1550

Qy        417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db       1551 GATAGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1610

Qy        477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db       1611 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1670

Qy        537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db       1671 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1730

Qy        597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db       1731 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 1790

Qy        657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db       1791 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 1850

Qy        717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db       1851 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 1910
```

```
Qy        777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db       1911 GGAGATTTGCACAGAGATGGAAAAGGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCC 1970

Qy        837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db       1971 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 2030

Qy        897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db       2031 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2090

Qy        957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db       2091 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2150

Qy       1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db       2151 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2210

Qy       1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db       2211 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2270

Qy       1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db       2271 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2330

Qy       1197 CCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250
Db       2331 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2390

Qy       1251 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1310
Db       2391 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2450

Qy       1311 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCA 1364
Db       2451 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 2510

Qy       1365 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1424
Db       2511 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2570

Qy       1425 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1484
Db       2571 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2630

Qy       1485 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1544
Db       2631 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2690

Qy       1545 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1604
Db       2691 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 2750

Qy       1605 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1664
Db       2751 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 2810

Qy       1665 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1724
Db       2811 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 2870

Qy       1725 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1784
Db       2871 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 2930

Qy       1785 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1844
Db       2931 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 2990

Qy       1845 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1904
Db       2991 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3050

Qy       1905 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1964
Db       3051 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3110

Qy       1965 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2024
Db       3111 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3170

Qy       2025 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2084
Db       3171 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3230

Qy       2085 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2144
Db       3231 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3290

Qy       2145 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2204
Db       3291 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3350

Qy       2205 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2264
Db       3351 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3410

Qy       2265 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2324
Db       3411 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3470

Qy       2325 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2357
Db       3471 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3503
```

RESULT 11
US-09-936-572-12
; Sequence 12, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 4642
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP2-codon
; OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence
US-09-936-572-12

```
  Query Match             65.1%;  Score 1599.4;  DB 4;  Length 4642;
  Best Local Similarity   81.5%;  Pred. No. 9.1e-264;
  Matches 1934;  Conservative    0;  Mismatches  411;  Indels   28;  Gaps    6;

Qy         12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68
Db       1421 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 1480

Qy         69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128
Db       1481 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 1540
```

```
Qy      129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188
Db     1541 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCA 1600

Qy      189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db     1601 AATGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTAGGGAAGATCTGGCCTTCCCACA 1659

Qy      249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db     1660 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGA 1719

Qy      309 GCTGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362
Db     1720 GCTTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGG 1779

Qy      363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db     1780 AACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAA 1839

Qy      417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db     1840 GATAGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 1899

Qy      477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db     1900 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1959

Qy      537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db     1960 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 2019

Qy      597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db     2020 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 2079

Qy      657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db     2080 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 2139

Qy      717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db     2140 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 2199

Qy      777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db     2200 GGAGATTTGCACAGAGATGGAAAAGGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCC 2259

Qy      837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db     2260 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 2319

Qy      897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db     2320 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2379

Qy      957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db     2380 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2439

Qy     1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db     2440 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 2499

Qy     1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db     2500 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 2559

Qy     1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db     2560 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 2619

Qy     1197 CCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250
Db     2620 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2679

Qy     1251 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1310
Db     2680 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 2739

Qy     1311 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCA 1364
Db     2740 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 2799

Qy     1365 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1424
Db     2800 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 2859

Qy     1425 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1484
Db     2860 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 2919

Qy     1485 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1544
Db     2920 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2979

Qy     1545 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1604
Db     2980 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 3039

Qy     1605 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1664
Db     3040 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 3099

Qy     1665 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1724
Db     3100 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 3159

Qy     1725 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1784
Db     3160 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 3219

Qy     1785 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1844
Db     3220 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 3279

Qy     1845 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1904
Db     3280 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 3339

Qy     1905 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1964
Db     3340 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 3399

Qy     1965 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2024
Db     3400 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 3459

Qy     2025 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2084
Db     3460 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 3519

Qy     2085 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2144
Db     3520 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 3579

Qy     2145 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2204
Db     3580 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 3639

Qy     2205 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2264
Db     3640 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 3699

Qy     2265 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2324
```

```
Db       3700 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 3759

Qy       2325 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2357

Db       3760 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 3792

RESULT 12
US-09-552-950-5
; Sequence 5, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
;  APPLICANT: Oxford Biomedica (UK) Limited
;  TITLE OF INVENTION: Anti-Viral Vectors
;  FILE REFERENCE: 674524-2004
;  CURRENT APPLICATION NUMBER: US/09/552,950
;  CURRENT FILING DATE:  2000-04-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 9772
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:pSYNGP
US-09-552-950-5

  Query Match             65.1%;  Score 1599.4;  DB 4;  Length 9772;
  Best Local Similarity   81.5%;  Pred. No. 1e-263;
  Matches 1934;  Conservative    0;  Mismatches  411;  Indels   28;  Gaps    6;

Qy         12 CATGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAA 68

Db       2193 CCTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCAGCGCGGCAA 2252

Qy         69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGC 128

Db       2253 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAAGAAGGGCACACAGC 2312

Qy        129 CCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCA 188

Db       2313 CCGCAACTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCA 2372

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248

Db       2373 AATGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACA 2431

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308

Db       2432 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGA 2491

Qy        309 GCTGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGG 362

Db       2492 GCTTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGG 2551

Qy        363 CA------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416

Db       2552 AACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAA 2611

Qy        417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476

Db       2612 GATAGGGGGGCAGCTCAAGGAGGCTCTCCTGGACACCGGAGCAGACGACACCGTGCTGGA 2671

Qy        477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536

Db       2672 GGAGATGTCGTTGCCAGGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 2731

Qy        537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596

Db       2732 CAAGGTGCGCCAGTATGACCAGATCCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 2791

Qy        597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656

Db       2792 CGTGCTGGTGGGCCCCACACCCGTCAACATCATCGGACGCAACCTGTTGACGCAGATCGG 2851

Qy        657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716

Db       2852 TTGCACGCTGAACTTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 2911

Qy        717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776

Db       2912 GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 2971

Qy        777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836

Db       2972 GGAGATTTGCACAGAGATGGAAAAGGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCC 3031

Qy        837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896

Db       3032 GTACAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT 3091

Qy        897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956

Db       3092 GGACTTCCGCGAGCTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 3151

Qy        957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016

Db       3152 GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 3211

Qy       1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076

Db       3212 CTTCTCCGTTCCCCTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGAT 3271

Qy       1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136

Db       3272 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGG 3331

Qy       1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196

Db       3332 CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAA 3391

Qy       1197 CCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250

Db       3392 CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT 3451

Qy       1251 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1310

Db       3452 AGGGCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGAC 3511

Qy       1311 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCA 1364

Db       3512 CACACCCGACAAGAAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCA 3571

Qy       1365 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1424

Db       3572 CCCTGACAAATGGACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA 3631

Qy       1425 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1484

Db       3632 CGACATACAGAAGCTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAA 3691

Qy       1485 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1544

Db       3692 GGTGAGGCAGCTGTGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 3751

Qy       1545 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1604

Db       3752 AACCGAGGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCA 3811

Qy       1605 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1664

Db       3812 CGGCGTGTACTATGACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG 3871

Qy       1665 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1724
```

```
Db     3872 CCAGTGGACCTATCAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 3931

Qy     1725 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1784
Db     3932 CCGGATGAGGGGTGCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGAT 3991

Qy     1785 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1844
Db     3992 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 4051

Qy     1845 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1904
Db     4052 GGAAACCTGGGAAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGA 4111

Qy     1905 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1964
Db     4112 GTTCGTCAACACCCCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGT 4171

Qy     1965 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2024
Db     4172 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGC 4231

Qy     2025 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2084
Db     4232 CGGATACGTCACTAACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCA 4291

Qy     2085 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2144
Db     4292 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACAT 4351

Qy     2145 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2204
Db     4352 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTC 4411

Qy     2205 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2264
Db     4412 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTG 4471

Qy     2265 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2324
Db     4472 GGTACCCGCCCACAAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGG 4531

Qy     2325 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGA 2357
Db     4532 CATCAGGAAGGTGCTATTCCTGGATGGCATCGA 4564
```

RESULT 13
US-09-872-733A-6
; Sequence 6, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 8366
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence
; OTHER INFORMATION: of the construct pCMVgagpolBNKan containing a CMV
; OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin
; OTHER INFORMATION: resistance gene

US-09-872-733A-6

```
  Query Match            62.6%;  Score 1539.2;  DB 4;  Length 8366;
  Best Local Similarity  79.9%;  Pred. No. 1.8e-253;
  Matches 1896;  Conservative    0;  Mismatches  448;  Indels   28;  Gaps    6;

Qy       14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70
Db     1857 TGGCCGAGGCGATGAGCCAGGTGACGAACTCGGCGACCATAATGATGCAGAGAGGCAACT 1916

Qy       71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db     1917 TCCGGAACCAGCGGAAGATCGTCAAGTGCTTCAATTGTGGCAAAGAAGGGCACACCGCCA 1976

Qy      131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db     1977 GGAACTGCCGGGCCCCCCGGAAGAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAA 2036

Qy      191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db     2037 TGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCTACAAG 2095

Qy      251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db     2096 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 2155

Qy      311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db     2156 TTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAA 2215

Qy      365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db     2216 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 2275

Qy      419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db     2276 TCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAG 2335

Qy      479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db     2336 AAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGGATCGGGGGCTTCATCA 2395

Qy      539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
Db     2396 AGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 2455

Qy      599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
Db     2456 TATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATCGGCT 2515

Qy      659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
Db     2516 GCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCGGGGA 2575

Qy      719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
Db     2576 TGGACGGCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCTTAGTCG 2635

Qy      779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
Db     2636 AAATCTGTACAGAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCT 2695

Qy      839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
Db     2696 ACAACACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGG 2755

Qy      899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
Db     2756 ACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCAC 2815

Qy      959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
Db     2816 ATCCCGCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACT 2875
```

```
Qy       1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
Db       2876 TCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCA 2935

Qy       1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
Db       2936 ACAACGAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGAT 2995

Qy       1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db       2996 CACCAGCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACC 3055

Qy       1199 CCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCG 1252
Db       3056 CAGACATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCG 3115

Qy       1253 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1312
Db       3116 GGCAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCA 3175

Qy       1313 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACC 1366
Db       3176 CACCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGCTACGAACTGCATC 3235

Qy       1367 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1426
Db       3236 CTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACG 3295

Qy       1427 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1486
Db       3296 ACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAG 3355

Qy       1487 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1546
Db       3356 TTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCCACTGA 3415

Qy       1547 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1606
Db       3416 CAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGTACATG 3475

Qy       1607 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1666
Db       3476 GAGTGTACTACGACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCAAGGCC 3535

Qy       1667 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1726
Db       3536 AATGGACCTACCAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTACGCAA 3595

Qy       1727 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1786
Db       3596 GGATGAGGGGTGCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAAGATCA 3655

Qy       1787 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1846
Db       3656 CCACAGAGAGCATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACAGAAGG 3715

Qy       1847 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1906
Db       3716 AGACATGGGAGACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGGAGT 3775

Qy       1907 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1966
Db       3776 TCGTGAACACCCCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGG 3835

Qy       1967 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2026
Db       3836 GAGCAGAGACCTTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAAGGCAG 3895

Qy       2027 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2086
Db       3896 GCTACGTGACCAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAACCAGA 3955

Qy       2087 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2146
Db       3956 AGACTGAGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCG 4015

Qy       2147 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2206
Db       4016 TGACAGACTCACAGTACGCACTGGGCATCATCCAAGCACAACCAGACCAATCCGAGTCAG 4075

Qy       2207 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2266
Db       4076 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGCATGGG 4135

Qy       2267 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2326
Db       4136 TACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGGA 4195

Qy       2327 TCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db       4196 TCCGGAAGGTGCTGTTCCTGGACGGGATCGAT 4227
```

RESULT 14
US-09-872-733A-1
; Sequence 1, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4338
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
;   OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene
US-09-872-733A-1

```
  Query Match             61.2%;  Score 1503.8;  DB 4;  Length 4338;
  Best Local Similarity   79.8%;  Pred. No. 1.7e-247;
  Matches 1896;  Conservative    0;  Mismatches  447;  Indels   34;  Gaps    9;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAACT 70
Db       1085 TGGCCGAGGCGATGAGCCAGGTGACGAACTCGGCGACCATAATGATGCAGAGAGGCAACT 1144

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db       1145 TCCGGAACCAGCGGAAGATCGTCAAGTGCTTCAATTGTGGCAAAGAAGGGCACACCGCCA 1204

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       1205 GGAACTGCCGGGCCCCCCGGAAGAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAA 1264

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       1265 TGAAAGATTGTACTGAGAGACAGGCTAA-TTTTTAGGGAAGATCTGGCCTTCCTACAAG 1323

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       1324 GGAAGGCCAGGGAATTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 1383

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
```

```
Db     1384 TTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAA 1443

Qy      365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418

Db     1444 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGA 1503

Qy      419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478

Db     1504 TCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAG 1563

Qy      479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538

Db     1564 AAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGGATCGGGGGCTTCATCA 1623

Qy      539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598

Db     1624 AGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 1683

Qy      599 TGCTGATCGGCCCCACC----CCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654

Db     1684 TATTAGTAGGACCTACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATC 1743

Qy      655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714

Db     1744 GGCTGCACCTTGAACTTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCG 1803

Qy      715 GGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTG 774

Db     1804 GGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCTTA 1863

Qy      775 ACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAAC 834

Db     1864 GTCGAAATCTGTACAGAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAAC 1923

Qy      835 CCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTG 894

Db     1924 CCCTACAACACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTG 1983

Qy      895 GTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954

Db     1984 GTGGACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATC 2043

Qy      955 CCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCC 1014

Db     2044 CCACATCCCGCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCC 2103

Qy     1015 TACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGC 1074

Db     2104 TACTTCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGC 2163

Qy     1075 ATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAG 1134

Db     2164 ATCAACAACGAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAG 2223

Qy     1135 GGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGC 1194

Db     2224 GGATCACCAGCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAA 2283

Qy     1195 AACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAG 1248

Db     2284 AACCCAGACATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAG 2343

Qy     1249 ATC-GGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTT 1307

Db     2344 ATCGGGGCAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACT 2403

Qy     1308 CACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCT 1361

Db     2404 GACCACACCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACT 2463

Qy     1362 GCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGT 1421

Db     2464 GCATCCTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGT 2523

Qy     1422 GAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCAT 1481

Db     2524 GAACGACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCAT 2583

Qy     1482 CAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCC 1541

Db     2584 CAAAGTTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCC 2643

Qy     1542 CCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGT 1601

Db     2644 ACTGACAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGT 2703

Qy     1602 GCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCA 1661

Db     2704 ACATGGAGTGTACTACGACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCA 2763

Qy     1662 CGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTA 1721

Db     2764 AGGCCAATGGACCTACCAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTA 2823

Qy     1722 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1781

Db     2824 CGCAAGGATGAGGGGTGCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAA 2883

Qy     1782 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1841

Db     2884 GATCACCACAGAGAGCATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACA 2943

Qy     1842 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1901

Db     2944 GAAGGAGACATGGGAGACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTG 3003

Qy     1902 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1961

Db     3004 GGAGTTCGTGAACACCCCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCAT 3063

Qy     1962 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2021

Db     3064 CGTGGGAGCAGAGACCTTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAA 3123

Qy     2022 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2081

Db     3124 GGCAGGCTACGTGACCAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAA 3183

Qy     2082 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2141

Db     3184 CCAGAAGACTGAGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAA 3243

Qy     2142 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2201

Db     3244 CATCGTGACAGACTCACAGTACG-CATGGGCATCATCCAAGCACAACCAGACCAATCCGA 3302

Qy     2202 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2261

Db     3303 GTCAGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGC 3362

Qy     2262 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2321

Db     3363 ATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGC 3422

Qy     2322 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358

Db     3423 TGGGATCCGGAAGGTGCTGTTCCTGGACGGGATCGAT 3459
```

RESULT 15
US-09-952-060-1
; Sequence 1, Application US/09952060
; Patent No. 6733993
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.
; APPLICANT: Youil, Rima

```
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION: MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2577
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Codon optimized DNA encoding modified wt pol
US-09-952-060-1

  Query Match           52.2%;  Score 1282.8;  DB 4;  Length 2577;
  Best Local Similarity 85.6%;  Pred. No. 7.3e-210;
  Matches 1454;  Conservative    0;  Mismatches  232;  Indels   12;  Gaps    2;

Qy        672 CCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAA 731
Db         15 CCCCATCTCCCCCATTGAGACTGTGCCTGTGAAGCTGAAGCCTGGCATGGATGGCCCCAA 74

Qy        732 GGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGA 791
Db         75 GGTGAAGCAGTGGCCCCTGACTGAGGAGAAGATCAAGGCCCTGGTGGAAATCTGCACTGA 134

Qy        792 GATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGT 851
Db        135 GATGGAGAAGGAGGGCAAAATCTCCAAGATTGGCCCCGAGAACCCCTACAACACCCCTGT 194

Qy        852 GTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCT 911
Db        195 GTTTGCCATCAAGAAGAAGGACTCCACCAAGTGGAGGAAGCTGGTGGACTTCAGGGAGCT 254

Qy        912 GAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCT 971
Db        255 GAACAAGAGGACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCTGGCCT 314

Qy        972 GAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCT 1031
Db        315 GAAGAAGAAGAAGTCTGTGACTGTGCTGGATGTGGGGGATGCCTACTTCTCTGTGCCCCT 374

Qy       1032 GGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCC 1091
Db        375 GGATGAGGACTTCAGGAAGTACACTGCCTTCACCATCCCCTCCATCAACAATGAGACCCC 434

Qy       1092 CGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTT 1151
Db        435 TGGCATCAGGTACCAGTACAATGTGCTGCCCCAGGGCTGGAAGGGCTCCCCTGCCATCTT 494

Qy       1152 CCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGAT 1211
Db        495 CCAGTCCTCCATGACCAAGATCCTGGAGCCCTTCAGGAAGCAGAACCCTGACATTGTGAT 554

Qy       1212 CTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGC 1265
Db        555 CTACCAGTACATGGATGACCTGTATGTGGGCTCTGACCTGGAGATTGGGCAGCACAGGAC 614

Qy       1266 CAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAA 1325
Db        615 CAAGATTGAGGAGCTGAGGCAGCACCTGCTGAGGTGGGGCCTGACCACCCCTGACAAGAA 674

Qy       1326 GCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGAC 1379
Db        675 GCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTATGAGCTGCACCCCGACAAGTGGAC 734

Qy       1380 CGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCT 1439
Db        735 TGTGCAGCCCATTGTGCTGCCTGAGAAGGACTCCTGGACTGTGAATGACATCCAGAAGCT 794

Qy       1440 GGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTG 1499
Db        795 GGTGGGCAAGCTGAACTGGGCCTCCCAAATCTACCCTGGCATCAAGGTGAGGCAGCTGTG 854

Qy       1500 CAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGA 1559
Db        855 CAAGCTGCTGAGGGGCACCAAGGCCCTGACTGAGGTGATCCCCCTGACTGAGGAGGCTGA 914

Qy       1560 GCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGA 1619
Db        915 GCTGGAGCTGGCTGAGAACAGGGAGATCCTGAAGGAGCCTGTGCATGGGGTGTACTATGA 974

Qy       1620 CCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCA 1679
Db        975 CCCCTCCAAGGACCTGATTGCTGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCA 1034

Qy       1680 GATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGC 1739
Db       1035 AATCTACCAGGAGCCCTTCAAGAACCTGAAGACTGGCAAGTATGCCAGGATGAGGGGGGC 1094

Qy       1740 CCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCAT 1799
Db       1095 CCACACCAATGATGTGAAGCAGCTGACTGAGGCTGTGCAGAAGATCACCACTGAGTCCAT 1154

Qy       1800 CGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGAC 1859
Db       1155 TGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGACCTGGGAGAC 1214

Qy       1860 CTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCC 1919
Db       1215 CTGGTGGACTGAGTACTGGCAGGCCACCTGGATCCCTGAGTGGGAGTTTGTGAACACCCC 1274

Qy       1920 CCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTT 1979
Db       1275 CCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATTGTGGGGGCTGAGACCTT 1334

Qy       1980 CTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGA 2039
Db       1335 CTATGTGGATGGGGCTGCCAACAGGGAGACCAAGCTGGGCAAGGCTGGCTATGTGACCAA 1394

Qy       2040 CCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCA 2099
Db       1395 CAGGGGCAGGCAGAAGGTGGTGACCCTGACTGACACCACCAACCAGAAGACTGAGCTCCA 1454

Qy       2100 GGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCA 2159
Db       1455 GGCCATCTACCTGGCCCTCCAGGACTCTGGCCTGGAGGTGAACATTGTGACTGACTCCCA 1514

Qy       2160 GTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCA 2219
Db       1515 GTATGCCCTGGGCATCATCCAGGCCCAGCCTGATCAGTCTGAGTCTGAGCTGGTGAACCA 1574

Qy       2220 GATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAA 2279
Db       1575 GATCATTGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGCCTGCCCACAA 1634

Qy       2280 GGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCT 2339
Db       1635 GGGCATTGGGGGCAATGAGCAGGTGGACAAGCTGGTGTCTGCTGGCATCAGGAAGGTGCT 1694
```

```
Qy       2340 GTTCCTGGACGGCATCGA 2357
              ||||||||| ||||| ||
Db       1695 GTTCCTGGATGGCATTGA 1712
```

Search completed: June 2, 2005, 03:13:16
Job time : 399.052 secs

```
                       GenCore version 5.1.6
              Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         June  1, 2005, 07:40:05 ; Search time 7836.86 Seconds
                                           (without alignments)
                                           11933.846 Million cell updates/sec

Title:          US-09-610-313B-32
Perfect score:  2457
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2457

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters:      68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_htc:*
                4:  gb_est3:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_gss1:*
                9:  gb_gss2:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                    %
Result             Query
  No.    Score    Match Length DB  ID                   Description
-------------------------------------------------------------------------------
c  1     136.4     5.6     330  9  CL293849             CL293849 02S0349-0
   2       100     4.1    2886  9  CL967755             CL967755 OsIFCC015
   3      90.2     3.7    2598  3  AY103647             AY103647 Zea mays
   4      88.8     3.6    3069  9  CL973991             CL973991 OsIFCC025
   5      86.6     3.5    1398  9  CL961989             CL961989 OsIFCC006
   6      85.6     3.5     951  4  BM321451             BM321451 rockefell
   7      84.2     3.4    1941  9  CL971508             CL971508 OsIFCC021
c  8      80.6     3.3     869  7  CK159167             CK159167 FGAS04056
   9      80.2     3.3    1132  4  BM320864             BM320864 rockefell
  10      79.6     3.2    1165  4  BM320900             BM320900 rockefell
  11      79.6     3.2    1680  9  CL982770             CL982770 OsIFSC049
  12        79     3.2     867  4  BM321430             BM321430 rockefell
  13      78.4     3.2    1725  9  CL978463             CL978463 OsIFCC031
  14      78.4     3.2    2031  9  CL974989             CL974989 OsIFCC042
  15      78.2     3.2    2697  9  CL952258             CL952258 OsJRUA000
  16      78.2     3.2    2853  9  CL974397             CL974397 OsIFCC025
  17        78     3.2    1485  9  CL970981             CL970981 OsIFCC020
  18      77.4     3.2    1509  9  CL959255             CL959255 OsIFCC002
  19      77.4     3.2    2559  9  CL982027             CL982027 OsIFSC046
  20        77     3.1   11691  9  CL962901             CL962901 OsIFCC008
  21      76.8     3.1    2682  9  CL969033             CL969033 OsIFCC017
  22      76.6     3.1    1550  4  BM321022             BM321022 rockefell
  23      76.2     3.1    1401  9  CL962721             CL962721 OsIFCC038
  24      76.2     3.1    3249  9  CL945510             CL945510 OsIFSB004
  25        75     3.1     545  4  BI724851             BI724851 1031075E0
  26        75     3.1     862  4  BM321023             BM321023 rockefell
  27      74.4     3.0     914  8  BZ568300             BZ568300 pacs2-164
  28      74.2     3.0    2313  9  CL982362             CL982362 OsIFSC047
  29        74     3.0    1290  9  CL972679             CL972679 OsIFCC023
  30        74     3.0    2072  3  CR603312             CR603312 full-leng
  31      73.8     3.0    2433  9  AY401196             AY401196 Homo sapi
  32      73.2     3.0     853  4  BM321393             BM321393 rockefell
c 33      72.8     3.0     566  4  BM587428             BM587428 170006873
  34      72.6     3.0     788  6  CB643171             CB643171 OSJNEb03L
  35      72.6     3.0     849  3  CR718981             CR718981 Tetraodon
  36      72.4     2.9     753  9  CC675888             CC675888 OGWCO51TH
  37      72.4     2.9    2523  9  CL974879             CL974879 OsIFCC026
  38      72.2     2.9     602  7  CV057146             CV057146 BNEL24e3
  39      72.2     2.9     640  2  BE601575             BE601575 HVSMEh009
  40      72.2     2.9     688  6  CB648640             CB648640 OSJNEb12C
  41      72.2     2.9     764  6  CB651670             CB651670 OSJNEb16L
  42      72.2     2.9     766  6  CB642928             CB642928 OSJNEb03F
  43      72.2     2.9     809  6  CB641397             CB641397 OSJNEb01A
  44      72.2     2.9     841  6  CB651502             CB651502 OSJNEb16H
c 45      72.2     2.9     851  9  CG260623             CG260623 OGWGE95TV

                                ALIGNMENTS

RESULT 1
CL293849/c
LOCUS       CL293849                 330 bp    DNA     linear   GSS 12-FEB-2004
DEFINITION  02S0349-08A1-C03 UniformMu MuTAIL Library Zea mays genomic clone
            02S0349-08A1-C03, genomic survey sequence.
ACCESSION   CL293849
VERSION     CL293849.1  GI:42541978
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Latshaw,S., Tan,B-.C., Settles,A.M. and McCarty,D.R.
  TITLE     Sequence tagged transposon insertions from the UniformMu maize
            population
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Donald R. McCarty
            Plant Molecular and Cellular Biology Program
            University of Florida
            PO 110690 Gainesville, FL 32611-0690, USA
            Tel: 352-392-1928 x322
            Email: drm@ufl.edu
            Sequence flanking probable Mu insertion site in UniformMu line:
            02S0349-08, Primer set: A
            Class: transposon insertion site.
FEATURES             Location/Qualifiers
     source          1..330
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="W22 (ACR, bz1-m9)"
                     /cultivar="UniformMu"
                     /db_xref="taxon:4577"
                     /clone="02S0349-08A1-C03"
                     /clone_lib="UniformMu MuTAIL Library"
                     /note="Vector: TOPO-PCR4; DNA flanking Mu transposon
                     insertions in Mu inactive lines were extracted from the
                     UniformMu maize population by the thermo assymmetric
                     interlaced PCR (TAIL) protocol using primers specific for
                     the Mu terminal inverted repeat and a set of 16 arbitrary
                     primers. Amplicons were size enriched using Sepharose 400
                     spin columns and cloned into the TOPO PCR4 vector."
ORIGIN

  Query Match             5.6%;  Score 136.4;  DB 9;  Length 330;
  Best Local Similarity   66.4%;  Pred. No. 2.5e-14;
```

```
  Matches  211;  Conservative    0;  Mismatches  106;  Indels    1;  Gaps    1;

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Db        329 GGGGGGCAGCTGGAAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAA 270

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Db        269 ATGACTTTGACAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAA 210

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Db        209 GTAAGACAGTATGATCAGGTACCCATAGAAATCTGTGGGCATAAAGCTATAGGTACGGTA 150

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Db        149 TTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGG-TGC 91

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Db         90 ACCTTAAATTTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATG 31

Qy        721 GACGGCCCCAAGGTGAAG 738

Db         30 GATGGCCCAAAAGTAAGG 13


RESULT 2
CL967755
LOCUS       CL967755                2886 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC015718 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL967755
VERSION     CL967755.1  GI:52390149
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2886)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..2886
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             4.1%;  Score 100;  DB 9;  Length 2886;
  Best Local Similarity   42.8%;  Pred. No. 9.6e-08;
  Matches  700;  Conservative    0;  Mismatches  915;  Indels   21;  Gaps    3;

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248

Db        219 GGTCATGGAGGCCGCCGCCATCATGGCCATCGTCCTCGCCAACGGCGGCGGCAGGCCGCC 278

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308

Db        279 CGACTGGCAGGACTTCGTCGGCATCGTCGTCCTCCTCGTCATCAACTCCACCATCAGTTT 338

Qy        309 GCTGCAGGTGCGC---GGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCAC 365

Db        339 CATCGAGGAGAACAATGCCGGGAACGCCGCCGCGGCGCTCATGGCCGGCCTCGCGCCCAA 398

Qy        366 CCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGG 425

Db        399 GACCAAGGTGCTCAGGGACGGGAAATGGCAGGAGCAGGACGCGTCCATCCTCGTTCCCGG 458

Qy        426 CCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAG 485

Db        459 CGACATCATCAGCATCAAGCTCGGCGACATCATCCCCGCCGACGCGCGGCTGCTCGAGGG 518

Qy        486 CCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCG 545

Db        519 CGACCCGCTCAAGGTCGACCAGGCGGCGCTCACCGGCGAGTCGATGCCCGTCAACAAGCA 578

Qy        546 CCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGAT 605

Db        579 CGCCGGGCAGGGCGTCTTCTCCGGGTCCACCGTCAAGCAGGGCGAGATCGAGGCCGTCGT 638

Qy        606 CGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCT 665

Db        639 CATCGCCACCGGCGTGCACACCTTCTTCGGCAAGGCGGCGCACCTGGTGGACAGCACCAA 698

Qy        666 GAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGG 725

Db        699 CAACATCGGCCACTTCCAGCTCGTGCTCACGGCCATCGGCAACTTCTGCATCATCTCCAT 758

Qy        726 CCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTG 785

Db        759 CGGCGTCGGCATGATAATCGAGATCATCGTTATGTACCCGATCCAGCACCGCGCGTACCG 818

Qy        786 CGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACAC 845

Db        819 CGACGGCATCGACAACCTGCTCGTCCTGCTCATCGGCGGCATCCCCATCGCCATGCCCAC 878

Qy        846 CCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCG 905

Db        879 CGTGCTGTCCGTCACCATGGCCATCGGCTCCCACCGGCTGTCGCAGCAGGGCGCCATCAC 938

Qy        906 CGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGC 965

Db        939 CAAGCGGATGACCGCCATCGAGGAGATGGCCGGCATGGACGTGCTCTGCAGCGACAAGAC 998

Qy        966 CGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGT 1025

Db        999 CGGGACGCTGACGCTCAACAAGCTCACCGTCGACAAGACGCTGATCGAGGTGTACGGGCG 1058

Qy       1026 GCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGA 1085

Db       1059 GGGGCTGGACAAGGACTCGGTGCTCCTGTACGCGGCGAGGGCGTCCCGCGTCGAGAACCA 1118

Qy       1086 GACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAG 1145

Db       1119 GGACGCCATTGACACATGCATCGTGGGCATGCTCGCCGACCCCAAGGAGGCCCGCGCCGG 1178

Qy       1146 CATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGAT 1205

Db       1179 CATCAAGGAGGTCCACTTCCTC------------CCGTTCAACCCCGTCGAGAAGCGCAC 1226

Qy       1206 CGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGC 1265

Db       1227 GGCCATCACCTACATCGACGGCAATGGCGAATGGCACAGGATCAGCAAGGGCGCGCCGGA 1286

Qy       1266 CAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAA 1325

Db       1287 GCAGATCATCGAGCTGTGCAAGATGAGCAAGGACGCCGAGAAGAAGGTGCACACGCTGAT 1346

Qy       1326 GCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCA 1385
```

```
Db     1347 CGACCAGTACGCGGACCGCGGCCTCCGGTCGCTGGGCGTGTCGTACCAGAAGGTGCCGGA 1406

Qy     1386 GCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGG 1445

Db     1407 GAAGAGCAAGGAGAGCGAGGGCGAGCCATGGCAGTTCGTCGGCCTCCTGCCGCTGTTCGA 1466

Qy     1446 CAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCT 1505

Db     1467 CCCGCCGAGGCACGACAGCGCGGAGACCATCCGCCGCGCGCTGCACCTCGGCGTGAACGT 1526

Qy     1506 GCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGA 1565

Db     1527 GAAGATGATCACCGGCGACCAGCTCGCCATCG------GGAAGGAGACGGCGCGGCGGCT 1580

Qy     1566 GCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAG 1625

Db     1581 CGGGATGGGCACCAACATGTACCCGTCGACGACGCTGCTGGGCGACAAGAGCAGCGAGAT 1640

Qy     1626 CAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTA 1685

Db     1641 GAGCGGCCTCCCAATCGACGAGCTGATCGAGAAGGCCGACGGCTTCGCCGGCGTGTTCCC 1700

Qy     1686 CCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACAC 1745

Db     1701 GGAGCACAAGTACGAGATCGTGAAGCGCCTCCAGGACCGGAAGCACATCTGCGGCATGAC 1760

Qy     1746 CAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGAT 1805

Db     1761 CGGCGACGGCGTGAACGACGCGCCGGCGCTGAAGAAGGCCGACATCGGCATCGCCGTCGA 1820

Qy     1806 CTGGGGCAAGACCCCC 1821

Db     1821 CGACGCCACGGACGCC 1836
```

RESULT 3
AY103647

```
LOCUS       AY103647              2598 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO142084 mRNA sequence.
ACCESSION   AY103647
VERSION     AY103647.1  GI:21206725
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2598)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2598)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..2598
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:638378"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

  Query Match              3.7%;  Score 90.2;  DB 3;  Length 2598;
  Best Local Similarity   42.2%;  Pred. No. 5.4e-06;
  Matches  814;  Conservative    0;  Mismatches 1098;  Indels   15;  Gaps    5;

Qy      296 CCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGC 355

Db       51 CCACCACCTCCCACCTCCTGCCACCGCGCGGCCCCAACCAACCACACCGCAGCGCAGCAA 110

Qy      356 GCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCA 415

Db      111 TGGCGGGCAACGGCGCCATCGTGGAGAGCGACCCGCTGAACTGGGGCGCGGCGGCGGCGG 170

Qy      416 AGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGG 475

Db      171 AGCTGGCCGGGAGCCACCTGGACGAGGTGAAGCGCATGGTGGCGCAGGCCCGGCAGCCCG 230

Qy      476 AGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCA 535

Db      231 TGGTCAAGATCGAGGGTTCCACCCTCCGCGTCGGCCAGGTGGCCGCCGTCGCCTCCGCCA 290

Qy      536 TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCA 595

Db      291 AGGACGCGTCCGGCGTCGCCGTCGAGCTCGACGAGGAGGCCCGCCCCCGCGTCAAGGCCA 350

Qy      596 CCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGG 655

Db      351 GCAGCGAGTGGATCCTCGACTGCATCGCCCACGGCGGCGACATCTACGGCGTCACCACCG 410

Qy      656 GCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCG 715

Db      411 GCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGGCCCGCGCTCCAGGTCGAGCTGC 470

Qy      716 GCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGA 775

Db      471 TCAGGCATCTCAACGCCGGAATCTTCGGCACCGGCAGCGACGGGCACACGCTGCCGTCGG 530

Qy      776 CCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACC 835

Db      531 AGGTCACCCGCGCGGCGATGCTGGTGCGCATCAACACCCTCCTCCAGGGCTACTCCGGCA 590

Qy      836 CCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGG 895

Db      591 TCCGCTTCGAGATCCTCGAGGCCATCACGAAGCTGCTCAACACCGGTGTCAGCCCCTGCC 650

Qy      896 TGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCC 955

Db      651 TGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTGGTCCCGCTCTCCTACATCGCCG 710

Qy      956 CCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCT 1015

Db      711 GCCTCATCACGGGCCGCCCCAACGCGCAGGCCGTCACCG---TCGACGGGAGGAAGGTGG 767

Qy     1016 ACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCA 1075

Db      768 ACGCCGCCGAGGCGTTCAAGATCGCCGGCATCGAGGGCGGCTTCTTCAAGCTCAACCCCA 827

Qy     1076 TCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGG 1135

Db      828 AGGAGGGCCTCGCCATCGTCAACGGCACGTCCGTGGGCTCCGCGCTCGCGGCCACCGTGA 887

Qy     1136 GCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCA 1195

Db      888 TGTACGACGCCAACGTCCTGGCCGTCCTGTCGGAGGTCCT---GTCCGCCGTCTTCTGCG 944
```

```
Qy     1196 ACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255

Db      945 AGGTCATGAACGGCAAGCCCGAGTACACGGACCACCTGACCCACAAGCTGAAGCACCACC 1004

Qy     1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG---CGCTGGGGCTTCACCA 1312

Db     1005 CGGGGTCCATCGAGGCCGCGGCCATCATGGAGCACATCCTGGATGGCAGCTCCTTCATGA 1064

Qy     1313 CCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACA 1372

Db     1065 AGCAGGCCAAGAAGTTGAACGAGCTGGACCCGCTGCTGAAGCCCAAGCAGGACAGGTACG 1124

Qy     1373 AGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCC 1432

Db     1125 CGCTCCGCACGTCGCCGCAGTGGCTGGGCCCCCAGATCGAGGTCATCCGCGCCGCCACCA 1184

Qy     1433 AGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCC 1492

Db     1185 AGTCCATCGAGCGCGAGGTCAACTCCGTGAACGA---CAACCCGGTCATCGACGTCCACC 1241

Qy     1493 AGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGG 1552

Db     1242 GCGGCAAGGCGCTGCACGGCGGCAACTTCCAGGGCACCCCCATCGGCGTGTCCATGGACA 1301

Qy     1553 AGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGT 1612

Db     1302 ACGCCCGCCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCGAGCTCG 1361

Qy     1613 ACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGA 1672

Db     1362 TCAACGAGTTCTACAACAACGGGCTCACCTCCAACCTGGCCGGCAGCCGCAACCCCAGCC 1421

Qy     1673 CCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGC 1732

Db     1422 TGGACTACGGCTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCCGAGCTCCAGT 1481

Qy     1733 GCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGG 1792

Db     1482 ACCTGGGCAACCCCATCACCAACCACGTGCAGAGCGCGGAGCAGCACAACCAGGACGTGA 1541

Qy     1793 AGAGCAT---CGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1849

Db     1542 ACTCCCTGGGCCTCGTCTCGGCCAGGAAGACCGCCGAGGCGATCGACATCCTGAAGCTCA 1601

Qy     1850 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1909

Db     1602 TGTCGTCCACCTACATCGTGGCGCTGTGCCAGGCCGTGGACCTGCGCCACCTCGAGGAGA 1661

Qy     1910 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1969

Db     1662 ACATCAAGGCGTCGGTGAAGAACACCGTGACCCAGGTGGCCAAGAAGGTGCTGACCATGA 1721

Qy     1970 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029

Db     1722 ACCCCTCGGGCGAGCTCTCCAGCGCCCGCTTCAGCGAGAAGGAGCTGATCAGCGCCATCG 1781

Qy     2030 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2089

Db     1782 ACCGCGAGGCCGTGTTCACGTACGCGGAGGACGCGGCCAGCGGCAGCCTGCCGCTGATGC 1841

Qy     2090 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2149

Db     1842 AGAAGCTGCGCGCCGTGCTGGTGGACCACGCCCTCAGCAGCGGCGACGCGGAGCGGGAGC 1901

Qy     2150 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2209

Db     1902 CCTCCGTGTTCTCCAAGATCACCAGGTTCGAGGAGGAGCTCCGCGCGGTGCTGCCCCAGG 1961

Qy     2210 TGGTGAA 2216

Db     1962 AGGTGGA 1968
```

```
RESULT 4
CL973991
LOCUS       CL973991                3069 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC025416 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL973991
VERSION     CL973991.1  GI:52402507
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 3069)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..3069
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.6%;  Score 88.8;  DB 9;  Length 3069;
  Best Local Similarity   41.5%;  Pred. No. 9.8e-06;
  Matches  652;  Conservative    0;  Mismatches  917;  Indels    3;  Gaps    1;

Qy       37 ACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAG 96

Db      625 AGCGGCATCGGCAAGACGACGCTGGCGAGCTACCTCTACCGGAGGAACAGCATCCGGAAG 684

Qy       97 TGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAG 156

Db      685 CACTTCGCCTGCACGGCGTGGGTCCACGTCCCCCGCAAGTTCCGCTTCGCCGACCTCCTC 744

Qy      157 GGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCC 216

Db      745 CACGACGTCATCCGCCAGGCCAACGCCTCCTCCGGCGACGACGAGCTCGGCGGCGAGGC- 803

Qy      217 AACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAG 276

Db      804 --CGCCGACGACGACGTGCGGCGCCGGGGTGGGATGGAGGCGGAGAGGCTCGTCAAGGCG 861

Qy      277 CAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGC 336

Db      862 CGCCTCGCCGCCGCGCTCCAGGGGAAGAGGTACCTCGTGGTGCTCGACGACGTCCGGAGC 921

Qy      337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAG 396

Db      922 AGGGAGGAGTGGGCCTTCTTCCTCGCCGCGCTCCCCGGCGGCCTCGCCGGCAGCTGCGTG 981

Qy      397 CGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGC 456

Db      982 CTCGTCACGACGCAGGTCAAGATCAGCGAGGAGTCGGAGCAACCCGTGGCCGCCGGCGGC 1041

Qy      457 GCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATC 516

Db     1042 GGCGTCGTGCGCACCAGGGAGCTCGGGAAGCTGCGGCCGGAGCAGGCCGCCGAGCTGTTC 1101
```

```
Qy        517 GGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGC 576
Db       1102 CGCCGGAGGGTGTACGGCCACGGCGAGCCGGAGCCGAAGAAGATGGAGCAGCTCAAGAGC 1161

Qy        577 GGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGC 636
Db       1162 CTCGTGGACTCCATGACGAAGGGCTCCAACCTGCCGCTCAACATCGTCATGCTCGCCGGA 1221

Qy        637 AACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTG 696
Db       1222 CTCTTGAGGTCCAAGAAGGAGGATGAATGGGAGGCCGTGATCCATAGCCTCGATGATACG 1281

Qy        697 CCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAG 756
Db       1282 CCACCCCCGACATCGCCGGAGCAGACGCCGCCGGAGGCGGAGCCGAAGAAGGACAGCCTC 1341

Qy        757 GAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACC 816
Db       1342 GATGATAGCGAGACGGCGCCGACATCGACGGAGCAAACGAAGAAGACGAAGAAGAGGAAG 1401

Qy        817 AAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGC 876
Db       1402 CAGATGGCGACATCGATGGACAAGATACTGACGGTGTGCATGGACGACCTGCCGACGCAT 1461

Qy        877 ACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936
Db       1462 CTGAAGCCGTGCTTCCTCTACTTCGCCGGGTTCACCGCGCAGACGCCGATCTGCGCCGGG 1521

Qy        937 GAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTG 996
Db       1522 AAGCTGGTCCGGCTGTGGGTCGCCGAGGGGTTCATGCAGGCGAAGAACGGGCAGACGGTG 1581

Qy        997 CTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACC 1056
Db       1582 GAGGAGCACGGCGAGGAGTGCCTCAAGGAGCTCATCTCCCGGTGCCTCGTCCAGCTCGTC 1641

Qy       1057 GCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTG 1116
Db       1642 GAGACGGACGCCGGCGGCGGCAGGGTCGCCGCCGTCAGCATCCACCAAGCGGTGCTCGAC 1701

Qy       1117 CTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTG 1176
Db       1702 TTCGTCCAGGCGGAGGCGCGCGACACCAACTTCCTCCACGTGCACAGCGGCGCGGCGGGC 1761

Qy       1177 GAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGC 1236
Db       1762 CTCTCCAACGGCGCCGCCCGCCGCCTCGCGCTCCGCAACACCTACGACACCGACCTCGCC 1821

Qy       1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
Db       1822 GTCATGCTCGAGGCGCCCAAGCTCCACACCCTCCTCTGCGACATCCCGGAGCGCGCCGCC 1881

Qy       1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATC 1356
Db       1882 GCCGACGCCGCCGGCGCCGCGGCGTGGCAGAGAGCGCTCGAGCTCATCAATGGCCGCGCC 1941

Qy       1357 GAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGG 1416
Db       1942 CCCACCTTCTCCGTCCACGGCTCGAGGTTCCTCCGCGTGATGGATCTCAAGGGCGTGCGG 2001

Qy       1417 ACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCC 1476
Db       2002 CTGCCGCACCGGGAGTCGCTGCCGGAGGAGATCGGGTGGCTGATCCACCTCAGGTACCTG 2061

Qy       1477 GGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATC 1536
Db       2062 GGCCTCTCGCACACCACCATGAGGCACCTCCCGAGCTCCGTCAAGAGGCTCCGGAACCTG 2121

Qy       1537 GTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAG 1596
Db       2122 CAGACGCTGGACGTGTCGCGCACGGACGTCGAGGCGCTGCCGTGGCGGCTCTGGCGCAAC 2181

Qy       1597 CCCGTGCACGGC 1608
Db       2182 CCCTCGCTCCGC 2193
```

```
RESULT 5
CL961989
LOCUS       CL961989                1398 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC006991 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL961989
VERSION     CL961989.1  GI:52378720
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1398)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1398
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.5%;  Score 86.6;  DB 9;  Length 1398;
  Best Local Similarity   45.9%;  Pred. No. 2.3e-05;
  Matches  508;  Conservative    0;  Mismatches  569;  Indels   30;  Gaps    5;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db        166 GTCGAGAGGCACGTCGGCGAGGCCGACATCGTCTTCGTCTCCGTGAACACCCCCACCAAG 225

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db        226 GCCCGCGGCCTCGGCGCCGGCAAGGCCGCCG---ACCTCACCTACTGGGAGAGCGCGGCG 282

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db        283 CGGATGATCGCCGCCGTGGCGACGTCCGACAAGGTCGTCGTCGAGAAGTCCACCGTGCCG 342

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db        343 GTCAAGACCGCGGAGGCCATCGAGAAGATCCTCGACCACAACGGCCGCGACGGCGTCGGC 402

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db        403 TTCCAGATCCTCTCCAACCCGGAGTTCCTCGCCGAGGGCACCGCCATCCGCGACCTGCTC 462

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db        463 GCCCCCGA---------CCGCGTCCTCATCGGCGGCCGCGAGACCGCCGCGGGGCGCGCC 513

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
```

```
Db        514 GCCGTCCAGGCGCTCAAGGACGTGTACGCCCGATGGGTCCCCGAGGAGAGGATC---CTC 570

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db        571 ACCACCAACCTCTGGTCCGCCGAGCTGTCCAAGCTCGCCGCCAACGCGTTCCTGGCCCAG 630

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db        631 AGGATCTCGTCCGTCAACGCCATGTCCGCGCTCTGCGAGGCCACCGGCGCCGACGTCGCC 690

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGA---GATCTGCGGCAAGAAGGCCATCGGCACC 597
Db        691 GAGGTCGCCTACGCCGTGGGCAAGGACTCCAGGATCGGCGCCAAGTTTCTCAACGCCAGC 750

Qy        598 GTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGC 657
Db        751 GTGGGATTCGGTGGCTCCTGCTTCCAGAAAGACATCCTCAACCTGCTG------------ 798

Qy        658 TGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGC 717
Db        799 TACATCTGCGAGTGCAATGGCCTCCCCGAGGTGGCCAACTACTGGAAGCAGGTTATCAAG 858

Qy        718 ATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACC 777
Db        859 ATCAACGACTACCAGAAGAGCCGGTTCGTGAACCGCGTGGTGTCGTCCATGTTCAACACC 918

Qy        778 GCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCC 837
Db        919 GTGGCCGGCAAGAAGATCGCCGTGCTGGGCTTCGCCTTCAAGAAGGACACCGGCGACACC 978

Qy        838 TACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTG 897
Db        979 AGGGAGACGCCCGCCATCGACGTCTGCAAGGGGCTGATCGGCGACAAGGCCAAGGTGAGC 1038

Qy        898 GACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC 957
Db       1039 ATCTACGACCCCCAGGTGACGGAGGACCAGGTCCAGCGCGACCTCGCCATGAGCAAGTTC 1098

Qy        958 CACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTAC 1017
Db       1099 GATTGGGATCACCCGGTCCATCTCCAGCCGATGAGCCCCACGGCCATCAAGCAGGTGTCC 1158

Qy       1018 TTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATC 1077
Db       1159 GTCGCCTGGACGCCTACGAGGCGGCCAGGGACGCCCACGGCGTCTGCATCCTCACCGAG 1218

Qy       1078 AACAACGAGACCCCCGGCATCCGCTAC 1104
Db       1219 TGGGACGAGTTCAGGAGCCTCGACTAC 1245
```

```
RESULT 6
BM321451
LOCUS       BM321451                 951 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321451
VERSION     BM321451.1  GI:18055857
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 951)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
   MEDLINE  21819461
    PUBMED  11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 951   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .951
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match              3.5%;  Score 85.6;  DB 4;  Length 951;
  Best Local Similarity   46.0%;  Pred. No. 3.4e-05;
  Matches  323;  Conservative    0;  Mismatches  376;  Indels    3;  Gaps    1;

Qy       1722 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1781
Db        219 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 278

Qy       1782 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1841
Db        279 CATCGCGCAGGCCGGCGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 338

Qy       1842 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1901
Db        339 GGAGTGCACCTGGAAGGCCCTGTGCTTCGGCCCCTACCAGGGCCCTCAGATCATCGTCGA 398

Qy       1902 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1961
Db        399 CGACGGCGGTGACGCGACTCTGATGATCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 458

Qy       1962 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2021
Db        459 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 518

Qy       2022 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2081
Db        519 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 578

Qy       2082 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2141
Db        579 CGAGGAGACGACGACTGGCGTNATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 638

Qy       2142 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2201
Db        639 GTTCCCGGCCGTCAACGTCAACGAC---TCTNTCACCAAGAGCAAGTTCGACAACATCTA 695

Qy       2202 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2261
Db        696 CGGCTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGG 755

Qy       2262 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2321
Db        756 CAAGGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCG 815

Qy       2322 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
Db        816 CGGCCAGGGCTGCCGCGTCATCGTGACGGAGATCGACCCCATCTGCGCGCTGCAGGCGTC 875

Qy       2382 CATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGA 2423
Db        876 GATGGCCGGCTTCGAGGTCAACACGCTCGAGGCGGGGCTCGA 917

RESULT 7
CL971508
LOCUS       CL971508                1941 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC021485 Oryza sativa Express Library Oryza sativa (indica
```

```
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL971508
VERSION     CL971508.1  GI:52397596
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1941)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1941
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.4%;  Score 84.2;  DB 9;  Length 1941;
  Best Local Similarity   42.5%;  Pred. No. 6.3e-05;
  Matches  705;  Conservative    0;  Mismatches  933;  Indels   21;  Gaps    4;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db        163 GACGCCGCCAAGAACCAGGTCGCCATGAACCCAACCAACACCGTCTTCGATGCGAAGCGG 222

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db        223 TTGATTGGTAGGAGGTTCTCCGACCCGTCCGTGCAGAGTGACATGAAGCTATGGCCGTTC 282

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db        283 AAGGTGGTGCCTGGCCCCGGCGACAAGCCGATGATCGTCGTCCAGTACAAGGGCGAGGAG 342

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db        343 AAGCAGTTCGCCGCGGAGGAGATCTCGTCCATGGTGCTGATCAAGATGAGGGAGATCGCG 402

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db        403 GAGGCCTACCTCGGCTCGTCCATCAAGAACGCCGTGGTCACCGTCCCGGCCTACTTCAAC 462

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db        463 GACTCGCAGAGGCAGGCCACCAAGGACGCCGGCGTCATCGCCGGGCTCAACGTGATGCGC 522

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db        523 ATCATCAACGAGCCCACCGCCGCCGCCATCGCGTACGGCCTCGACAAGAAGGCGACGAGC 582

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db        583 AGCGGCGAGAAGAACGTGCTCATCTTTGACCTCGGCGGCGGCACGTTCGATGTGTCGCTG 642

Qy        481 ATGAGCCT---GCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATC 537
Db        643 CTCACCATCGAGGAGGGCATCTTCGAGGTCAAGGCCACCGCCGGCGACACCCACCTCGGC 702

Qy        538 AAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG---AAGGCCATCGGC 594
Db        703 GGCGAGGACTTCGACAACCGCATGGTGAACCACTTCGTGCAGGAGTTCAAGCGCAAGAAC 762

Qy        595 ACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654
Db        763 AAGAAGGACATCAGCGGCAACCCGCGCGCGCTGCGGCGGCTCAGGACGGCGTGCGAGCGC 822

Qy        655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714
Db        823 GCCAAGCGCACGCTGTCGTCGACGGCGCAGACCACCATTGAGATCGACTCGCTGTACGAG 882

Qy        715 GGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTG 774
Db        883 GGCATCGACTTCTACACGACGATCACCAGGGCGCGGTTCGAGGAGCTCAACATGGACCTG 942

Qy        775 ACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAAC 834
Db        943 TTCCGCAAGTGCATGGAGCCCGTGGAGAAGTGCCTCCGCGACGCCAAGATGGACAAGAGC 1002

Qy        835 CCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTG 894
Db       1003 AGCGTGCACGACGTCGTGCTCGTCGGCGGCTCCACCCGCATCCCCAAGGTGCAGCAGCTG 1062

Qy        895 ------------GTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
Db       1063 CTCCAGGACTTCTTCAACGGCAAGGAGCTCTGCAAGAGCATCAACCCCGACGAGGCCGTG 1122

Qy        943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
Db       1123 GCGTACGGCGCCGCCGTCCAGGCCGCCATCCTGAGCGGCGAGGGC---AACGAGAAGGTG 1179

Qy       1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
Db       1180 CAGGACCTGCTCCTGCTCGACGTCACCCCGCTGTCCCTCGGCCTCGAGACCGCCGGCGGC 1239

Qy       1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
Db       1240 GTCATGACCGTGCTCATCCCGAGGAACACCACCATCCCGACCAAGAAGGAGCAGGTCTTC 1299

Qy       1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
Db       1300 TCCACCTACTCCGACAACCAGCCCGGCGTCCTGATCCAGGTGTACGAGGGCGAGCGCGCC 1359

Qy       1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGAC 1242
Db       1360 AGGACCAAGGACAACAACCTCCTCGGCAAGTTCGAGCTCTCCGGCATCCCACCGGCGCCG 1419

Qy       1243 CTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGG 1302
Db       1420 CGCGGCGTGCCCCAGATCAACGTCTGCTTCGACATCGACGCCAACGGCATCCTGAACGTG 1479

Qy       1303 GGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTG 1362
Db       1480 TCCGCGGAGGACAAGACGACGGGGCAGAAGAACAAGATCACCATCACCAACGACAAGGGG 1539

Qy       1363 CACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTG 1422
Db       1540 CGGCTGAGCAAGGAGGAGATCGAGAAGATGGTGCAGGAGGCGGAGAAGTACAAGGCCGAG 1599

Qy       1423 AACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATC 1482
Db       1600 GACGAGGAGCACAAGAAGAAGGTGGACGCCAAGAACGCCCTCGAGAACTACGCCTACAAC 1659

Qy       1483 AAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCC 1542
Db       1660 ATGCGCAACACCATCAAGGACGAGAAGATCGCCTCCAAGCTCGCCGCCGACGACAAGAAG 1719

Qy       1543 CTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTG 1602
Db       1720 AGGATCGAGGACGCCATCGATGGCGCCATCAGCTGGCTCGACACCAACCAGCTCGCCGAG 1779
```

```
Qy       1603 CACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCC 1641

Db       1780 GCCGACGAGTTCGAGGACAAGATGAAGGAGCTGGAGGGC 1818


RESULT 8
CK159167/c
LOCUS       CK159167                869 bp    mRNA    linear   EST 05-DEC-2003
DEFINITION  FGAS040564 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA,
            mRNA sequence.
ACCESSION   CK159167
VERSION     CK159167.1  GI:38985053
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
            Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
            Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
            Penniket,C., Roach,J.L. and Sarhan,F.
  TITLE     Functional Genomics of Abiotic Stress In Wheat and Canola Crops
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [128,636].
            Plate: TaLt537  row: N  column: 23.
FEATURES             Location/Qualifiers
     source          1..869
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Wheat line PI 178383"
                     /db_xref="taxon:4565"
                     /lab_host="DH5 alpha"
                     /clone_lib="Triticum aestivum FGAS: TaLt5"
                     /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
                     subtractive hybridization) cDNA library from genotype
                     PI178383 cold hardened at 2 C for 21 days and 49 days
                     (equal amount of cDNA pooled together before subtraction,
                     tester) and subtracted against genotype Norstar cold
                     hardened at 2 C for 1 day (24 H)(driver). Modified Smart
                     cDNA (Clontech)priming and non-directional cloning"
ORIGIN

  Query Match             3.3%;  Score 80.6;  DB 7;  Length 869;
  Best Local Similarity   44.9%;  Pred. No. 0.00027;
  Matches  305;  Conservative    0;  Mismatches  374;  Indels    0;  Gaps    0;

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639

Db        827 ATGGCGGCGGCCCGCACAACGAGGATCACCAACCAGGACGAGAACAACCACAACCACCAC 768

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699

Db        767 CACAACGACGACAACAGCGACACAACGACCACCAACAACACCAACAACAACAACAACAAC 708

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759

Db        707 ACCAACAACAACAACAACAACAACAACAACAACGACAACCACCACCACAACCACAACAAC 648

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819

Db        647 AACAACAACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 588

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879

Db        587 AACAACAACAACAACAACAACAACAACAACAACAACGACAACAACAACAACAACAACAAC 528

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939

Db        527 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 468

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999

Db        467 AACAACAACAACAACAACAACAACAACGACGACGACGACAACGACGACGACGACGACGAC 408

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059

Db        407 GACAACAACGACAACAACGACAACAACAACAACGACAACAACAACAACAACAACAACAAC 348

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119

Db        347 GACCACAACGACAACGACAACAACAACAACAACAACAACAACAACGACAACAACCACAAC 288

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179

Db        287 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 228

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGC 1239

Db        227 AACAACAACAACAACAACGACGACAACGACGTCAACGATAACAACGACAACAACGACAAC 168

Qy       1240 GACCTGGAGATCGGCCAGC 1258

Db        167 AACGGCAACAACAACAAAC 149


RESULT 9
BM320864
LOCUS       BM320864               1132 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM320864
VERSION     BM320864.1  GI:18055270
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1132)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
   MEDLINE  21819461
    PUBMED  11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1132   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..1132
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN
```

```
  Query Match             3.3%;  Score 80.2;  DB 4;  Length 1132;
  Best Local Similarity   45.3%;  Pred. No. 0.00032;
  Matches  375;  Conservative    0;  Mismatches  444;  Indels    9;  Gaps    2;

Qy        139 CGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGAC 198
Db         79 CGCCGCGAGGGCAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAG 138

Qy        199 TGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCC 258
Db        139 TACAACAGCCCCAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTC 192

Qy        259 CGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG 318
Db        193 TGCCAGATCGCCTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCAC 252

Qy        319 CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCC 378
Db        253 GAGCTCACCCGCTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACT 312

Qy        379 CAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAG 438
Db        313 GGCCTGCTGCTGGCCCGCCGTGTGCTGAAGAAGCTCAACCTCGACTCCAAGTACGAGGGT 372

Qy        439 GCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAG 498
Db        373 GTCAAGAAGGTCAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGT 432

Qy        499 TGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAG 558
Db        433 TTCAAGGCCCTGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCC 492

Qy        559 ATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCC 618
Db        493 GCCCTCAAGGGCATGTGCGACGGCGGCGTCAACGTCCCC---CACAGCGAGACCCGCTTC 549

Qy        619 GTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATC 678
Db        550 GTCGGCTTCAACGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTC 609

Qy        679 AGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAG 738
Db        610 GGCGGCCACGTCGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGAC 669

Qy        739 CAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAG 798
Db        670 CGCCAGTTCTCGCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTAC 729

Qy        799 AAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCC 858
Db        730 ACCGAGGCCCACAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAG 789

Qy        859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918
Db        790 CCCGAGGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAG 849

Qy        919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db        850 AACCGCGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCC 897
```

```
RESULT 10
BM320900
LOCUS       BM320900              1165 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM320900
VERSION     BM320900.1  GI:18055306
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1165)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1165   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..1165
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.2%;  Score 79.6;  DB 4;  Length 1165;
  Best Local Similarity   45.1%;  Pred. No. 0.00041;
  Matches  428;  Conservative    0;  Mismatches  506;  Indels   16;  Gaps    3;

Qy         90 CATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCG 149
Db         46 CGTCAAGAACAAGGCGTACTTCAAGCGCTTCCAGACCCAGTTCCGTCGCCGCCGCGAGGG 105

Qy        150 CAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCG 209
Db        106 CAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAGTACAACAGCCC 165

Qy        210 CCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCC 269
Db        166 CAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTCTGCCAGATCGC 219

Qy        270 CAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAA 329
Db        220 CTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCACGAGCTCACCCG 279

Qy        330 CCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCT 389
Db        280 CTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACTGGCCTGCTGCT 339

Qy        390 GTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGA 449
Db        340 GGCCCGCCGTGTGCTGAAGAAGCTCAACCTCTACTCCAAGTACGAGGGTGTCAAGAAGGT 399

Qy        450 CACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAA 509
Db        400 CAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGTTCAAGGCCC-- 457

Qy        510 GATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGA 569
Db        458 --TGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCCGCCCTCAA 515

Qy        570 GATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCAT 629
Db        516 GGGCATGTGCGACGGCGGCGTCAACGTCCCCCACAGCGAGACCCGCTTCGTCGGCTTCAA 575

Qy        630 CGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGA 689
Db        576 CGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTCGGCGGCCACGT 635

Qy        690 GACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCT 749
```

```
Db      636 CGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGACCGCCAGTTCTC 695

Qy      750 GACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAA 809

Db      696 GCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTACACCGAGGCCCA 755

Qy      810 GATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAA 869

Db      756 ------CAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAGCCCGA 809

Qy      870 GGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGA 929

Db      810 GGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAGAACCG 869

Qy      930 CTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGT 989

Db      870 CGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCCAAGAGCCAGTAAATTCC 929

Qy      990 GACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGG 1039

Db      930 CGGTGCCCTGTACACCGCCATCTCCCGCTGCGGCGTTGCGCTGCTNGCGG 979


RESULT 11
CL982770
LOCUS       CL982770                1680 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFSC049024 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL982770
VERSION     CL982770.1  GI:52420015
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1680)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1680
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.2%;  Score 79.6;  DB 9;  Length 1680;
  Best Local Similarity   41.9%;  Pred. No. 0.00042;
  Matches  612;  Conservative    0;  Mismatches  844;  Indels    6;  Gaps    2;

Qy      270 CAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAA 329

Db      129 CATCAACGTGGACCCGGGCACCATGTCCCCGTTCCAGCACGGCGAAGTGTTCGTCACCGA 188

Qy      330 CCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCT 389

Db      189 CGACGGCGCCGAGACCGACCTCGACCTCGGCCACTACGAGCGTTTCATCGAAACGCGCAT 248

Qy      390 GTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGA 449

Db      249 GAAGAAGACCAACAACTTCACCACGGGCCGCATCTACCAGTCCGTGCTGGAGAAGGAGCG 308

Qy      450 CACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAA 509

Db      309 CCGCGGCGACTACCTCGGCAAGACGGTGCAGGTGATCCCGCACGTCACCAACGAGATCCA 368

Qy      510 GATGATCGGC---GGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGAT 566

Db      369 GGAATACGTCAAGCGCGGCGCGGGCATCGGCACCGACGACGCGGTGGACGTGGCCATCGT 428

Qy      567 CGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACAT 626

Db      429 CGAGATCGGCGGCACCGTGGGCGACATCGAATCGCTGCCGTTCCTCGAAGCCGTGCGCCA 488

Qy      627 CATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCAT 686

Db      489 GATGAGCCTGCGCATGGGGCCGAACAATTCCGCCTTCGTGCACCTCACCTACCTGCCCTA 548

Qy      687 CGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCC 746

Db      549 CATCGCGGCGGCCGGTGAACTCAAGACCAAGCCCACCCAGCACACGGTGCAGAAGCTGCG 608

Qy      747 CCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGG 806

Db      609 CGAGATCGGCATCCAGCCCGACGCCCTGCTCTGCCGCGCCGACCGCCGCATCCCCGACGA 668

Qy      807 CAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAA 866

Db      669 GGAGCGCGGCAAGATCTCGCTGTTCACCAACGTGGCCGAATGGGGCGTGATCAGCATGTG 728

Qy      867 GAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCA 926

Db      729 GGACGTGGACACCATCTACAAGGTGCCCCGCATGCTGCACGAGCAGGGCCTGGACGGCCT 788

Qy      927 GGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAG 986

Db      789 GATCTGCGACAAGCTGCGCCTGAACACGCCGCCCACCAGCCTCAAGCGCTGGGACCAGCT 848

Qy      987 CGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCG 1046

Db      849 CGTGTACGAGACCGAGCATCCGCAGGGCGAGGTCACCATCGCCATGGTGGGCAAGTACGT 908

Qy     1047 CAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCA 1106

Db      909 AGACCTGTCCGACAGCTACAAGTCGGTGAACGAGGCGCTGCGCCACGCGGCATGAAGAA 968

Qy     1107 GTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGAC 1166

Db      969 CCACGTGCGCGTGAAGATCGAGCACGTCGATTCCGAGACCATCGGCGCGGCCGACGCGGC 1028

Qy     1167 CAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCT 1226

Db     1029 CGAGAAGCTGGCCAGGTACGACGCCATCCTGGTGCCCGGCGGCTTCGGCGCGCGCGGCGT 1088

Qy     1227 GTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAA 1286

Db     1089 GGAAGGCAAGATCTGCACGGCGCGCTACGCCCGCGAGCAGAAGGTGCCCTACCTGGGCAT 1148

Qy     1287 GCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTT 1346

Db     1149 CTGCCTGGGCATGCAGGTGGCCACCATCGAATATGCGCGCCACGTGGCGGGCCTCGCGAA 1208

Qy     1347 CCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAA 1406

Db     1209 CGCCAACAGCACCGAATTCGACGCCGCCACGCCGCATCCCGTGATCGCGCTGATCACCGA 1268

Qy     1407 GGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCA 1466

Db     1269 GTGGAAGGACGCCGACGGCACCATCCAGACGCGCAGCGAGACCTCCGACCTGGGCGGCAC 1328

Qy     1467 GATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCT 1526
```

```
Db       1329 CATGCGCCTGGGCGCGCAAAGCTCCGACGTCGCCTCCGGCACGCTGGCGCACGAGATCTA 1388

Qy       1527 GACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGAT 1586

Db       1389 CGGCAACATCGTCACCGAGCGCCACCGCCACCGGTACGAGGCGAACGTGAACTACCTCGA 1448

Qy       1587 CCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAG---GACCTGGTGGCCGA 1643

Db       1449 CCAGCTCCGCAAGGCGGGCCTCGTGATCTCCGCGCTCACGCAGCGCGAACAGCTCACCGA 1508

Qy       1644 GATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAA 1703

Db       1509 GATCGTCGAGCTGCCGCGCGAGGTGCATCCGTGGTTCATCGGCGTGCAGTTCCACCCCGA 1568

Qy       1704 CCTGAAGACCGGCAAGTACGCC 1725

Db       1569 GTTCAAGTCCACCCCGTGGGAC 1590
```

```
RESULT 12
BM321430
LOCUS       BM321430                867 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
            sequence.
ACCESSION   BM321430
VERSION     BM321430.1  GI:18055836
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 867   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..867
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.2%;  Score 79;  DB 4;  Length 867;
  Best Local Similarity   45.7%;  Pred. No. 0.00051;
  Matches  314;  Conservative    0;  Mismatches  370;  Indels    3;  Gaps    1;

Qy         58 CAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAG 117

Db        131 CCGCACAAGATGCGCGAGTGCCTGCCGGTCATCATCCTGCTGCGCAACAGGTTGAAGTAC 190

Qy        118 GGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAG 177

Db        191 GCGCTGACCCGCCGTGAGGTCACCTCGATCGTGATGCAGCGCCTGATCAAGATCGACGGC 250

Qy        178 GAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTG 237

Db        251 AAGGTCCGCACCGACACGCACTTCCCCGCGGGCTTCATGGACGTCGTCTCGATCGACAAG 310

Qy        238 GCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCC 297

Db        311 ACCGACGAGCACTTCCGCCTGCTCTACGACACCAAGGGCCGCTTCCAGGCGCACCGCATC 370

Qy        298 ACCAGCCGCGAGCTGCAGGT---GCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAG 354

Db        371 AACTCGACCGAGGCCAAGTTCAAGCTCGGCAAGGTCCGCCGCGTGCAGCTCGGCAACAAG 430

Qy        355 CGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATC 414

Db        431 GGCATCCCGTACCTGGTGACCCACGACGCGCGCACGATCCGCTACCCCAACCCCGACATC 490

Qy        415 AAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTG 474

Db        491 AAGGTCAACGACACGGTCAAGATCGACCTGCGCTCGGGCAAGATCATCGACTTCGTCAGG 550

Qy        475 GAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTC 534

Db        551 TTCGAGATCGGCAACCTCGTCATGATCACTGGCGGACGCAACCTTGGCCGCGTCGGCGTC 610

Qy        535 ATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGC 594

Db        611 ATTGTGCGCCGCGAGAAGCACGAGGGCTCGTTCGAGATCATCCACGTCAAGGACGCCGTC 670

Qy        595 ACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654

Db        671 GGCCACCAGTTCGCGACGCGCCTGACCAACGTCTTCGTGATCGGCAAGGGCACCAAGTCC 730

Qy        655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714

Db        731 CTCGTCACGCTGCCCGCCGGCAAGGGCATCAAGAAGTCGATCATCGAGGAGTTCCAGGCG 790

Qy        715 GGCATGGACGGCCCCAAGGTGAAGCAG 741

Db        791 CGCCACGCGCACAGGACCGAGCAGGAG 817
```

```
RESULT 13
CL978463
LOCUS       CL978463               1725 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC031917 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL978463
VERSION     CL978463.1  GI:52411427
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1725)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1725
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
```

```
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.2%;  Score 78.4;  DB 9;  Length 1725;
  Best Local Similarity   43.5%;  Pred. No. 0.00069;
  Matches  468;  Conservative    0;  Mismatches  596;  Indels   12;  Gaps    2;

Qy        949 GGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGC 1008
Db         76 GTCCTCCGCACCCCGGACGCCCTCCACCGCATGTACGCGGAGGTGAGCCTCCTCGCCGAC 135

Qy       1009 GACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATC 1068
Db        136 CTCCGCCACGACGCCATCATCGCCTTCCACGCCTCCTGGGTCCACCCCTCCCGCCGCACC 195

Qy       1069 CCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGC 1128
Db        196 TTCAACTTCATCACCGAGCTCTTCTCCTCCGGCACCCTCCGCTCCTACCGCCTCCGCTAC 255

Qy       1129 TGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGC 1188
Db        256 CCGCGCGTCAGCCGCCGCGCCGTCGCCGCCTGGGCGCGCGCCATCCTCCACGGCCTCGCC 315

Qy       1189 GCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAG 1248
Db        316 TACCTCCACTCCCGCGGCGTCATCCACCGCGACCTCAAGTGCGACAACATCTTCGTCAAC 375

Qy       1249 ATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTC 1308
Db        376 GGCCACCTCGGCCAGGTCAAGATCGGCGACCTCGGCCTCGCCGCCGTCCTCCGCGGCTGC 435

Qy       1309 ACCACCCCCGACA---------AGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAG 1359
Db        436 ACCTCCGCCCGCAGCGTCATCGGCACGCCGGAGTTCATGGCCCCCGAGATGTACGACGAG 495

Qy       1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
Db        496 TGCTACGGCGTCGGCGTCGACGTCTACTCCTTCGGCATGTGCATGCTCGAGATGCTCACC 555

Qy       1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
Db        556 AACGAGTACCCCTACAGCGAGTGCGACAACCCGGCGCAGATCTACAAGAAGGTCACCGCC 615

Qy       1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG-- 1537
Db        616 GGCAAGCTCCCCGACGCCTTCTACCTCCTCACCGACGCCGACGCCCGCCGCTTCATCGGC 675

Qy       1538 -TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAG 1596
Db        676 CGGTGCCTTGTCGACGCCGCCCACCGCCCCTCTGCCGAGGAGCTCCTGCTCGACCCCTTC 735

Qy       1597 CCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAG 1656
Db        736 CTCTCCCCGCCTCAGAACCACGACGATCACAATACCATTGCTCATGCCACTGCGCCGCCA 795

Qy       1657 GGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGC 1716
Db        796 CCGCCATTGCCATTGGCGTGCAGTAATTCCAGTGAGGAGCAAGAGGAGGAGGAGGCGCCG 855

Qy       1717 AAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTG 1776
Db        856 GCAGCAAAGACGACGGGCATGGCGATCACCGGGAAGCTGAACAAGGAGCACGACACCATC 915

Qy       1777 CAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCC 1836
Db        916 TTCCTCAAGGTGCAGATCGGCGGCGGCGGCAACGTGCGGAACATCTACTTCCCGTTCGAC 975

Qy       1837 ATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCC 1896
Db        976 GTGGCGAACGACACGGCGATGGAGGTGGCGACGGAGATGGTGAAGGAGCTGGACATCGCC 1035

Qy       1897 GAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAG 1956
Db       1036 GACCGGGAGCCCACGGAGATTGCCGCCATGATCGAGCAGGAGATCGTGAGGTTGGTACCA 1095

Qy       1957 CCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAA 2012
Db       1096 GGGTACAAGCAGCACGAGTATTCGTACGCGGATGACGACGACGATGATGACGTCAA 1151


RESULT 14
CL974989
LOCUS       CL974989                2031 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC042895 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL974989
VERSION     CL974989.1  GI:52404497
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2031)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..2031
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.2%;  Score 78.4;  DB 9;  Length 2031;
  Best Local Similarity   42.1%;  Pred. No. 0.0007;
  Matches  719;  Conservative    0;  Mismatches  966;  Indels   21;  Gaps    4;

Qy        672 CCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAA 731
Db        285 CGCCAAGAACCAGGCGGCGGCCAACCCGGAGCGGACCATCTACGACGCCAAGCGGCTCAT 344

Qy        732 GGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGA 791
Db        345 CGGCCGGCAGTTCTCCGACGCCGAGGTGCAGCGCGACATGAAGCTGCTGCCGTTCGCCGT 404

Qy        792 GATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGT 851
Db        405 CGTCGACCGGAACGGCAAGCCGCACGTGCGCGTCGAGGTGAAGGACGGCGACGTGCGCGT 464

Qy        852 GTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCT 911
Db        465 GTTCAGCCCGGAGGAGGTGAGCGCCATGGTGCTCACGCGGATGAAGGAGACGGCCGAGGC 524

Qy        912 GAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCT 971
Db        525 CTACCTCGGCGAGAAGGTCACCCGCGCCGTCGTCACCGTCCCGGCCTACTTCAACGACGC 584

Qy        972 GAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCT 1031
Db        585 GCAGCGGCAGGCCACCAAGGACGCCGGCGTCATCGCCGGGCTCACCGTCGACCGCATCAT 644
```

```
Qy     1032 GGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCC 1091
Db      645 CAACGAGCCCACCGCCGCCGCCATCGCCTACGGCATCGACAAGAAGGGCGCCGAGAAGAA 704

Qy     1092 CGGCATCCGCTACCAGTACAACG------TGCTGCCCCAGGGCTGGAAGGGCAGCCCCAG 1145
Db      705 CGTCCTCGTCTTCGACCTCGGCGGCGGCACGTTCGACGTCAGCATCCTCGCCATCGACAA 764

Qy     1146 CATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGAT 1205
Db      765 CGGCGTGTTCGAGGTCCTTGCCACCAACGGCGACACCCACCTCGGCGGCGAGGACTTCGA 824

Qy     1206 CGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGC 1265
Db      825 CCAACGCCTCATGGACCACTTCGTCAAGGTCATCCGCCGGAAGCACGGCCGCGACATCAC 884

Qy     1266 CAAGATCGAG---GAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAA 1322
Db      885 CGGCGACGCGCGCGCGCTGGGCAAGCTCCGCCGCGAGTGCGAGCGCGCCAAGCGCGCGCT 944

Qy     1323 GAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGT 1382
Db      945 CAGCAACCAGCACCAGGTGCGCGTCGAGGTCGAGTCCCTGTTCGACGGCGTCGACTTGTC 1004

Qy     1383 GCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGT 1442
Db     1005 GGAGCCGCTCTCCCGGGCGCGGTTCGAGGAGCTCAACAGCGACCTCTTCAAGAAGACGAT 1064

Qy     1443 GGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAA 1502
Db     1065 GGTGCCGGTGAGGAAGGCCATGGCGGACGCCCGGCTGAGCAAGGGCGACATCGACGAGAT 1124

Qy     1503 GCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCT 1562
Db     1125 CGTCCTCGTCGGCGGCAGCACCAGGATCCCCAAGGTGCAGCAGCTGCTCAAGGACTACTT 1184

Qy     1563 GGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCC 1622
Db     1185 CGGTGGCAAGGAGCCCAACCGCGGCGTCAACCCCGACGAGGCCGTGGCGTACGGCGCCGC 1244

Qy     1623 CAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGAT 1682
Db     1245 CGTGCAGGCCAGCATCATCAGCGGCCACGTCGACGAGAACACGGAGAGCATGATCCTCCT 1304

Qy     1683 CTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCA 1742
Db     1305 CGACGTCGCGCCGCTCACCCTCGGCCTGGAGACGGCCGGCGGCGTGATGGCCAAGCTGAT 1364

Qy     1743 CACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGT 1802
Db     1365 CCCGCGCAACACGGTGGTGCCGACGAAGAAGACGCAGGTGTTCACCACGTACAAGGACAA 1424

Qy     1803 GATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTG 1862
Db     1425 GCAGACAACGGTGACCATCCAGGTGTTCGAGGGCGAGCGGAGCATGACGAGGGACAACCG 1484

Qy     1863 GTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCC 1922
Db     1485 GCTGCTCGGCAGGTTCGACCTCGCCGGCATCGCGCCGGCGCCGAGGGGG------GCGCC 1538

Qy     1923 CCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTA 1982
Db     1539 GCAGATCGAGGTGACGTTCGAGGTGGACGCGAACGGCATCCTCAGCGTGCTGGC------ 1592

Qy     1983 CGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCG 2042
Db     1593 CGCCGACAAGGCCACCGGCAGGTCGGAGAAGATCACCATCTCCGGCGACGACCGCAAGAT 1652

Qy     2043 GGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGC 2102
Db     1653 CAGCCAGGAGGAGATCGACCGGATGGTGCGCGAGGCGGAGGAGTTCGCCGAGGAGGACCG 1712

Qy     2103 CATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTA 2162
Db     1713 CCGGCACAGGGAGCAGGTGGACGCCCGGAACAGCCTGGAGGCGTACGTCTACAACATCAA 1772

Qy     2163 CGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGAT 2222
Db     1773 GAACACGCTCGGCGGCAAGATGGCGGACGCCATGGAGGGCGAGGAGAAGGACAAGGTGGA 1832

Qy     2223 CATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGG 2282
Db     1833 GGAGGCGGTGAGGGAGGCGTACGAGTGGCTGGACGGCAACCCGGACGCCGGCAAGGAGGA 1892

Qy     2283 CATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTT 2342
Db     1893 GTACGAGGAGAAGCTGAGGGAGCTGGAGGACGTGTGCAACCCCGTCATGTCGGCGGTCTA 1952

Qy     2343 CCTGGACGGCATCGATGGCGGCATCG 2368
Db     1953 CCAGAGGTCCGGCGGCGGCGGCGGCG 1978
```

```
RESULT 15
CL952258
LOCUS       CL952258              2697 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsJRUA000500 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL952258
VERSION     CL952258.1  GI:52364267
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2697)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..2697
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.2%;  Score 78.2;  DB 9;  Length 2697;
  Best Local Similarity   43.4%;  Pred. No. 0.00077;
  Matches  680;  Conservative    0;  Mismatches  863;  Indels   24;  Gaps    6;

Qy      193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
Db     1141 AAGGACTACGCGGTGAGGCTGAGCAACGACACCGCCGACGAGCATCTGTGGGTGGCGGTG 1200

Qy      253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
Db     1201 CACCCCTCGGTGATGCTCCGCCCGCAGTTCTACGACGCCATCCTCAACGGCCTCGAGGTG 1260

Qy      313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
Db     1261 TTCAAGGTGAACAACACCGGCGGCAGCCTGGCGTCGCCGGACCCCGTCCCGTACAAGCTG 1320
```

```
Qy        373 TTCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
Db       1321 CTCGCGGAGAAGGAGCTCGGCTGGGGCGGGCCGCCGGAGTTCTCCACCGACAACCCGGCC 1380

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
Db       1381 AACATGGCGTCGGTGATGGGCGGCACGGCGGGCGGCGCGGCGGCGGCCGGGATCGTGGCG 1440

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
Db       1441 GCCATCTG------CGTGGTGGTGTACAGCAACAAGAGGAGCAAGAAGCTGGGCGGCGGC 1494

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
Db       1495 GGCGCCGACTCGCACACCTCCGCGTGGCTGCCGCTGTACCACTCCCACACCAGCGGCAAG 1554

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
Db       1555 TCGTCGGGGCACATCACGGCGAACATCGCCGGCATGTGCCGCCACTTCTCGTTCGCGGAG 1614

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
Db       1615 ATCAAGGCGGCGACCAAGAACTTCTCCAACGACCTGGCCATCGGCGTGGGCGGGTTCGGC 1674

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
Db       1675 GTGGTGTACCGCGGCGTGGTGGACGGCGACGTGAAGG---TGGCGGTGAAGCGGTCCAAC 1731

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db       1732 CCGTCGTCGGAGCAAGGCATAACCGAGTTCCAGACGGAGGTGGAGATGCTCTCCAAGCTC 1791

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db       1792 CGCCACCGCCACCTCGTCTCCCTCATCGGCTTCTGCGAGGAGGACGGCGAGATGGTGCTC 1851

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db       1852 GTCTACGACTACATGGAGCACGGCACCCTGCGCGAGCACCTCTACCACAACGGCGGCAAG 1911

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db       1912 CCCACGCTGTCGTGGCGCCACCGCCTCGACATCTGCATCGGCGCCGCCCGCGGCCTCC-- 1969

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db       1970 ----ACTACCTCCACACCGGCGCCAAGTACACCATCATCCACCGCGACGTCAAGACCACC 2025

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db       2026 AACATCCTCGTCGACGACAACTGGGTCGCCAAGGTCTCCGACTTCGGCCTCTCCAAGTCC 2085

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db       2086 GGCCCCACCACGCTCAA---CCAGTCGCACGTCAGCACCGTCGTCAAGGGCAGCTTCGGC 2142

Qy       1213 TACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATC 1272
Db       2143 TACCTCGACCCGGAGTACTACCGCCGGCAGCAGCTCACCGACAAGTCCGACGTCTACTCC 2202

Qy       1273 GAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAG 1332
Db       2203 TTCGGCGTCGTGCTGTTCGAGGTGCTCATGGCGCGCCCGGCGCTCGACCCGGCGCTGCCG 2262

Qy       1333 AAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATC 1392
Db       2263 CGCGACCAGGTCAGCCTCGCCGACTACGCGCTCGCCTGCAAGCGCGGCGGCGCGCTGCCG 2322

Qy       1393 GAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTG 1452
Db       2323 GACGTCGTCGACCCGGCGATCAGGGACCAGATCGCGCCCGAGTGCC---TCGCCAAGTTC 2379

Qy       1453 AACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGC 1512
Db       2380 GCCGACACGGCGGAGAAGTGCCTCTCCGAGAACGGCACCGAGCGCCCCACCATGGGCGAC 2439

Qy       1513 GGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCC 1572
Db       2440 GTGCTCTGGAACCT---CGAGAGCGCGATGCACTTCCAGGACGCGTTCGACGCCGCCGCC 2496

Qy       1573 GAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGAC 1632
Db       2497 GGCCGCCCCGTCCCCGCGCTCGACGCCGCCGCCGGCAGCAGCAGCCACCTCGACGACGGG 2556

Qy       1633 CTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAG 1692
Db       2557 AGCACGGCCAGCATCAACACGCTGGCCACGTCGTCCACGTCGCACCCGCACGAGCCGTGC 2616

Qy       1693 CCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGAC 1752
Db       2617 GTCGACGTTGTCTTGGAGCCCGACGACGTCGTCGCGGAGCGCGCCACCTTCTCGCAGCTC 2676

Qy       1753 GTGAAGC 1759
Db       2677 GTCCAGC 2683
```

Search completed: June 2, 2005, 02:53:33
Job time : 7843.86 secs

```
OM nucleic - nucleic search, using sw model

Run on:         June  1, 2005, 07:40:05 ; Search time 7856 Seconds
                                          (without alignments)
                                          11933.846 Million cell updates/sec

Title:          US-09-610-313B-31
Perfect score:  2463
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2463

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters:      68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_htc:*
                4:  gb_est3:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_gss1:*
                9:  gb_gss2:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	136.4	5.5	330	9	CL293849	CL293849 02S0349-0
2	98.4	4.0	2886	9	CL967755	CL967755 OsIFCC015
3	86.8	3.5	2598	3	AY103647	AY103647 Zea mays
4	86.6	3.5	1398	9	CL961989	CL961989 OsIFCC006
5	85.6	3.5	951	4	BM321451	BM321451 rockefell
6	85.4	3.5	2031	9	CL974989	CL974989 OsIFCC042
7	84	3.4	1725	9	CL978463	CL978463 OsIFCC031
8	80.8	3.3	3069	9	CL973991	CL973991 OsIFCC025
c 9	80.6	3.3	869	7	CK159167	CK159167 FGAS04056
10	80.6	3.3	1941	9	CL971508	CL971508 OsIFCC021
11	80.2	3.3	1132	4	BM320864	BM320864 rockefell
12	79.6	3.2	1165	4	BM320900	BM320900 rockefell
13	79.6	3.2	11691	9	CL962901	CL962901 OsIFCC008
14	79	3.2	867	4	BM321430	BM321430 rockefell
15	78.6	3.2	2682	9	CL969033	CL969033 OsIFCC017
16	78.2	3.2	2853	9	CL974397	CL974397 OsIFCC025
17	78	3.2	1485	9	CL970981	CL970981 OsIFCC020
18	78	3.2	2313	9	CL982362	CL982362 OsIFSC047
19	78	3.2	2454	9	CL975440	CL975440 OsIFCC027
20	77.4	3.1	1509	9	CL959255	CL959255 OsIFCC002
21	77.4	3.1	2559	9	CL982027	CL982027 OsIFSC046
22	76.6	3.1	1550	4	BM321022	BM321022 rockefell
23	76.4	3.1	2028	9	CL979437	CL979437 OsIFCC033
24	76.2	3.1	1401	9	CL962721	CL962721 OsIFCC038
25	76	3.1	1680	9	CL982770	CL982770 OsIFSC049
26	76	3.1	2793	9	CL948321	CL948321 OsIFSB004
27	75	3.0	862	4	BM321023	BM321023 rockefell
28	75	3.0	1734	9	CL948553	CL948553 OsIFSB005
29	74.6	3.0	1410	9	CL947314	CL947314 OsIFSB004
30	74.6	3.0	1989	9	CL980973	CL980973 OsIFCC035
31	74.4	3.0	914	8	BZ568300	BZ568300 pacs2-164
32	74	3.0	1290	9	CL972679	CL972679 OsIFCC023
33	74	3.0	2072	3	CR603312	CR603312 full-leng
34	73.2	3.0	853	4	BM321393	BM321393 rockefell
35	73	3.0	2151	9	CL972100	CL972100 OsIFCC041
c 36	72.8	3.0	566	4	BM587428	BM587428 170006873
37	72.6	2.9	788	6	CB643171	CB643171 OSJNEb03L
38	72.4	2.9	753	9	CC675888	CC675888 OGWCO51TH
39	72.2	2.9	602	7	CV057146	CV057146 BNEL24e3
40	72.2	2.9	640	2	BE601575	BE601575 HVSMEh009
41	72.2	2.9	688	6	CB648640	CB648640 OSJNEb12C
42	72.2	2.9	764	6	CB651670	CB651670 OSJNEb16L
43	72.2	2.9	766	6	CB642928	CB642928 OSJNEb03F
44	72.2	2.9	809	6	CB641397	CB641397 OSJNEb01A
45	72.2	2.9	841	6	CB651502	CB651502 OSJNEb16H

ALIGNMENTS

```
RESULT 1
CL293849/c
LOCUS       CL293849                330 bp    DNA     linear   GSS 12-FEB-2004
DEFINITION  02S0349-08A1-C03 UniformMu MuTAIL Library Zea mays genomic clone
            02S0349-08A1-C03, genomic survey sequence.
ACCESSION   CL293849
VERSION     CL293849.1  GI:42541978
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Latshaw,S., Tan,B-.C., Settles,A.M. and McCarty,D.R.
  TITLE     Sequence tagged transposon insertions from the UniformMu maize
            population
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Donald R. McCarty
            Plant Molecular and Cellular Biology Program
            University of Florida
            PO 110690 Gainesville, FL 32611-0690, USA
            Tel: 352-392-1928 x322
            Email: drm@ufl.edu
            Sequence flanking probable Mu insertion site in UniformMu line:
            02S0349-08, Primer set: A
            Class: transposon insertion site.
FEATURES             Location/Qualifiers
     source          1..330
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="W22 (ACR, bz1-m9)"
                     /cultivar="UniformMu"
                     /db_xref="taxon:4577"
                     /clone="02S0349-08A1-C03"
                     /clone_lib="UniformMu MuTAIL Library"
                     /note="Vector: TOPO-PCR4; DNA flanking Mu transposon
                     insertions in Mu inactive lines were extracted from the
                     UniformMu maize population by the thermo assymmetric
                     interlaced PCR (TAIL) protocol using primers specific for
                     the Mu terminal inverted repeat and a set of 16 arbitrary
                     primers. Amplicons were size enriched using Sepharose 400
                     spin columns and cloned into the TOPO PCR4 vector."
ORIGIN

  Query Match             5.5%;  Score 136.4;  DB 9;  Length 330;
  Best Local Similarity   66.4%;  Pred. No. 2.2e-14;
```

```
  Matches  211;  Conservative    0;  Mismatches  106;  Indels    1;  Gaps    1;

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Db        329 GGGGGGCAGCTGGAAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAA 270

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Db        269 ATGACTTTGACAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAA 210

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Db        209 GTAAGACAGTATGATCAGGTACCCATAGAAATCTGTGGCATAAAGCTATAGGTACGGTA 150

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Db        149 TTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGG-TGC 91

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Db         90 ACCTTAAATTTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATG 31

Qy        721 GACGGCCCCAAGGTGAAG 738

Db         30 GATGGCCCAAAAGTAAGG 13


RESULT 2
CL967755
LOCUS       CL967755                2886 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC015718 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL967755
VERSION     CL967755.1  GI:52390149
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2886)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..2886
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             4.0%;  Score 98.4;  DB 9;  Length 2886;
  Best Local Similarity   43.2%;  Pred. No. 1.7e-07;
  Matches  709;  Conservative    0;  Mismatches  906;  Indels   27;  Gaps    4;

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248

Db        219 GGTCATGGAGGCCGCCGCCATCATGGCCATCGTCCTCGCCAACGGCGGCGGCAGGCCGCC 278

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308

Db        279 CGACTGGCAGGACTTCGTCGGCATCGTCGTCCTCCTCGTCATCAACTCCACCATCAGTTT 338

Qy        309 GCTGCAGGTGCGC---GGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCAC 365

Db        339 CATCGAGGAGAACAATGCCGGGAACGCCGCCGCGGCGCTCATGGCCGGCCTCGCGCCCAA 398

Qy        366 CCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGG 425

Db        399 GACCAAGGTGCTCAGGGACGGGAAATGGCAGGAGCAGGACGCGTCCATCCTCGTTCCCGG 458

Qy        426 CCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAG 485

Db        459 CGACATCATCAGCATCAAGCTCGGCGACATCATCCCCGCCGACGCGCGGCTGCTCGAGGG 518

Qy        486 CCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCG 545

Db        519 CGACCCGCTCAAGGTCGACCAGGCGGCGCTCACCGGCGAGTCGATGCCCGTCAACAAGCA 578

Qy        546 CCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGAT 605

Db        579 CGCCGGGCAGGGCGTCTTCTCCGGGTCCACCGTCAAGCAGGGCGAGATCGAGGCCGTCGT 638

Qy        606 CGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCT 665

Db        639 CATCGCCACCGGCGTGCACACCTTCTTCGGCAAGGCGGCGCACCTGGTGGACAGCACCAA 698

Qy        666 GAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGG 725

Db        699 CAACATCGGCCACTTCCAGCTCGTGCTCACGGCCATCGGCAACTTCTGCATCATCTCCAT 758

Qy        726 CCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTG 785

Db        759 CGGCGTCGGCATGATAATCGAGATCATCGTTATGTACCCGATCCAGCACCGCGCGTACCG 818

Qy        786 CGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACAC 845

Db        819 CGACGGCATCGACAACCTGCTCGTCCTGCTCATCGGCGGCATCCCCATCGCCATGCCCAC 878

Qy        846 CCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCG 905

Db        879 CGTGCTGTCCGTCACCATGGCCATCGGCTCCCACCGGCTGTCGCAGCAGGGCGCCATCAC 938

Qy        906 CGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGC 965

Db        939 CAAGCGGATGACCGCCATCGAGGAGATGGCCGGCATGGACGTGCTCTGCAGCGACAAGAC 998

Qy        966 CGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGT 1025

Db        999 CGGGACGCTGACGCTCAACAAGCTCACCGTCGACAAGACGCTGATCGAGGTGTACGGGCG 1058

Qy       1026 GCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGA 1085

Db       1059 GGGGCTGGACAAGGACTCGGTGCTCCTGTACGCGGCGAGGGCGTCCCGCGTCGAGAACCA 1118

Qy       1086 GACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAG 1145

Db       1119 GGACGCCATTGACACATGCATCGTGGGCATGCTCGCCGACCCCAAGGAGGCCCGCGCCGG 1178

Qy       1146 CATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGAT 1205

Db       1179 CATCAAGGAGGTCCACTTCCTCCCGTTC---AACCCCGTCGAGAAGCGCACGGCCATCAC 1235

Qy       1206 CGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGC 1265

Db       1236 CTACATCGACGGCAATGGCGAATGGCACAGGATCAGCAAGGGCGCGCCGGAGCAGATCAT 1295

Qy       1266 CAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAA 1325

Db       1296 CGAGCTGTGCAAGATGAGCAAGGACGCCGAGAAGAAGGTGCACACGCTGATCGACCAGTA 1355

Qy       1326 GCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGAC 1385
```

```
Db    1356 CGCGGACCGCGGCCTCCGGTCGCTGGGCGTGTCGTACCAGAAGGTGCCGGAGAAG----- 1410

Qy    1386 CGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCT 1445
Db    1411 ----------AGCAAGGAGAGCGAGGGCGAGCCATGGCAGTTCGTCGGCCTCCTGCCGCT 1460

Qy    1446 GGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTG 1505
Db    1461 GTTCGACCCGCCGAGGCACGACAGCGCGGAGACCATCCGCCGCGCGCTGCACCTCGGCGT 1520

Qy    1506 CAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGA 1565
Db    1521 GAACGTGAAGATGATCACCGGCGACCAGCTCGCCATCG------GGAAGGAGACGGCGCG 1574

Qy    1566 GCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGA 1625
Db    1575 GCGGCTCGGGATGGGCACCAACATGTACCCGTCGACGACGCTGCTGGGCGACAAGAGCAG 1634

Qy    1626 CCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCA 1685
Db    1635 CGAGATGAGCGGCCTCCCAATCGACGAGCTGATCGAGAAGGCCGACGGCTTCGCCGGCGT 1694

Qy    1686 GATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGC 1745
Db    1695 GTTCCCGGAGCACAAGTACGAGATCGTGAAGCGCCTCCAGGACCGGAAGCACATCTGCGG 1754

Qy    1746 CCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCAT 1805
Db    1755 CATGACCGGCGACGGCGTGAACGACGCGCCGGCGCTGAAGAAGGCCGACATCGGCATCGC 1814

Qy    1806 CGTGATCTGGGGCAAGACCCCC 1827
Db    1815 CGTCGACGACGCCACGGACGCC 1836
```

```
RESULT 3
AY103647
LOCUS       AY103647                2598 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO142084 mRNA sequence.
ACCESSION   AY103647
VERSION     AY103647.1  GI:21206725
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2598)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2598)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..2598
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:638378"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

  Query Match            3.5%;  Score 86.8;  DB 3;  Length 2598;
  Best Local Similarity  42.1%;  Pred. No. 2e-05;
  Matches  813;  Conservative    0;  Mismatches 1102;  Indels   15;  Gaps    5;

Qy     296 CCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGC 355
Db      51 CCACCACCTCCCACCTCCTGCCACCGCGCGGCCCCAACCAACCACACCGCAGCGCAGCAA 110

Qy     356 GCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCA 415
Db     111 TGGCGGGCAACGGCGCCATCGTGGAGAGCGACCCGCTGAACTGGGGCGCGGCGGCGGCGG 170

Qy     416 AGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGG 475
Db     171 AGCTGGCCGGGAGCCACCTGGACGAGGTGAAGCGCATGGTGGCGCAGGCCCGGCAGCCCG 230

Qy     476 AGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCA 535
Db     231 TGGTCAAGATCGAGGGTTCCACCCTCCGCGTCGGCCAGGTGGCCGCCGTCGCCTCCGCCA 290

Qy     536 TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCA 595
Db     291 AGGACGCGTCCGGCGTCGCCGTCGAGCTCGACGAGGAGGCCCGCCCCCGCGTCAAGGCCA 350

Qy     596 CCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGG 655
Db     351 GCAGCGAGTGGATCCTCGACTGCATCGCCCACGGCGGCGACATCTACGGCGTCACCACCG 410

Qy     656 GCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCG 715
Db     411 GCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGGCCCGCGCTCCAGGTCGAGCTGC 470

Qy     716 GCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGA 775
Db     471 TCAGGCATCTCAACGCCGGAATCTTCGGCACCGGCAGCGACGGGCACACGCTGCCGTCGG 530

Qy     776 CCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACC 835
Db     531 AGGTCACCCGCGCGGCGATGCTGGTGCGCATCAACACCCTCCTCCAGGGCTACTCCGGCA 590

Qy     836 CCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGG 895
Db     591 TCCGCTTCGAGATCCTCGAGGCCATCACGAAGCTGCTCAACACCGGTGTCAGCCCCTGCC 650

Qy     896 TGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCC 955
Db     651 TGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTGGTCCCGCTCTCCTACATCGCCG 710

Qy     956 CCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCT 1015
Db     711 GCCTCATCACGGGCCGCCCCAACGCGCAGGCCGTCACCG---TCGACGGGAGGAAGGTGG 767

Qy    1016 ACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCA 1075
Db     768 ACGCCGCCGAGGCGTTCAAGATCGCCGGCATCGAGGGCGGCTTCTTCAAGCTCAACCCCA 827

Qy    1076 TCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGG 1135
Db     828 AGGAGGGCCTCGCCATCGTCAACGGCACGTCCGTGGGCTCCGCGCTCGCGGCCACCGTGA 887

Qy    1136 GCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCA 1195
Db     888 TGTACGACGCCAACGTCCTGGCCGTCCTGTCGGAGGTCCT---GTCCGCCGTCTTCTGCG 944
```

```
Qy       1196 ACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255

Db        945 AGGTCATGAACGGCAAGCCCGAGTACACGGACCACCTGACCCACAAGCTGAAGCACCACC 1004

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315

Db       1005 CGGGGTCCATCGAGGCCGCGGCCATCATGGAGCACATCCTGGATGGCAGCTCCTTCATGA 1064

Qy       1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1375

Db       1065 AGCAGGCCAAGAAGTTGAACGAGCTGGACCCGCTGCTGAAGCCCAAGCAGGACAGGTACG 1124

Qy       1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1435

Db       1125 CGCTCCGCACGTCGCCGCAGTGGCTGGGCCCCCAGATCGAGGTC---ATCCGCGCCGCCA 1181

Qy       1436 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1495

Db       1182 CCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGA---CAACCCGGTCATCGACGTCC 1238

Qy       1496 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1555

Db       1239 ACCGCGGCAAGGCGCTGCACGGCGGCAACTTCCAGGGCACCCCCATCGGCGTGTCCATGG 1298

Qy       1556 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1615

Db       1299 ACAACGCCCGCCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCGAGC 1358

Qy       1616 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1675

Db       1359 TCGTCAACGAGTTCTACAACAACGGGCTCACCTCCAACCTGGCCGGCAGCCGCAACCCCA 1418

Qy       1676 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1735

Db       1419 GCCTGGACTACGGCTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCCGAGCTCC 1478

Qy       1736 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1795

Db       1479 AGTACCTGGGCAACCCCATCACCAACCACGTGCAGAGCGCGGAGCAGCACAACCAGGACG 1538

Qy       1796 TGGAGAGCAT---CGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1852

Db       1539 TGAACTCCCTGGGCCTCGTCTCGGCCAGGAAGACCGCCGAGGCGATCGACATCCTGAAGC 1598

Qy       1853 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1912

Db       1599 TCATGTCGTCCACCTACATCGTGGCGCTGTGCCAGGCCGTGGACCTGCGCCACCTCGAGG 1658

Qy       1913 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1972

Db       1659 AGAACATCAAGGCGTCGGTGAAGAACACCGTGACCCAGGTGGCCAAGAAGGTGCTGACCA 1718

Qy       1973 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2032

Db       1719 TGAACCCCTCGGGCGAGCTCTCCAGCGCCCGCTTCAGCGAGAAGGAGCTGATCAGCGCCA 1778

Qy       2033 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2092

Db       1779 TCGACCGCGAGGCCGTGTTCACGTACGCGGAGGACGCGGCCAGCGGCAGCCTGCCGCTGA 1838

Qy       2093 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2152

Db       1839 TGCAGAAGCTGCGCGCCGTGCTGGTGGACCACGCCCTCAGCAGCGGCGACGCGGAGCGGG 1898

Qy       2153 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2212

Db       1899 AGCCCTCCGTGTTCTCCAAGATCACCAGGTTCGAGGAGGAGCTCCGCGCGGTGCTGCCCC 1958

Qy       2213 AGCTGGTGAA 2222

Db       1959 AGGAGGTGGA 1968
```

```
RESULT 4
CL961989
LOCUS       CL961989                1398 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC006991 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL961989
VERSION     CL961989.1  GI:52378720
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1398)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1398
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.5%;  Score 86.6;  DB 9;  Length 1398;
  Best Local Similarity   45.9%;  Pred. No. 2.1e-05;
  Matches  508;  Conservative    0;  Mismatches  569;  Indels   30;  Gaps    5;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Db        166 GTCGAGAGGCACGTCGGCGAGGCCGACATCGTCTTCGTCTCCGTGAACACCCCCACCAAG 225

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Db        226 GCCCGCGGCCTCGGCGCCGGCAAGGCCGCCG---ACCTCACCTACTGGGAGAGCGCGGCG 282

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Db        283 CGGATGATCGCCGCCGTGGCGACGTCCGACAAGGTCGTCGTCGAGAAGTCCACCGTGCCG 342

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Db        343 GTCAAGACCGCGGAGGCCATCGAGAAGATCCTCGACCACAACGGCCGCGACGGCGTCGGC 402

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Db        403 TTCCAGATCCTCTCCAACCCGGAGTTCCTCGCCGAGGGCACCGCCATCCGCGACCTGCTC 462

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Db        463 GCCCCCGA---------CCGCGTCCTCATCGGCGGCCGCGAGACCGCCGCGGGGCGCGCC 513

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCCTGGTGAGCATCAAGGTG 420

Db        514 GCCGTCCAGGCGCTCAAGGACGTGTACGCCCGATGGGTCCCCGAGGAGAGGATC---CTC 570

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Db        571 ACCACCAACCTCTGGTCCGCCGAGCTGTCCAAGCTCGCCGCCAACGCGTTCCTGGCCCAG 630
```

```
Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Db        631 AGGATCTCGTCCGTCAACGCCATGTCCGCGCTCTGCGAGGCCACCGGCGCCGACGTCGCC 690

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGA---GATCTGCGGCAAGAAGGCCATCGGCACC 597

Db        691 GAGGTCGCCTACGCCGTGGGCAAGGACTCCAGGATCGGCGCCAAGTTTCTCAACGCCAGC 750

Qy        598 GTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGC 657

Db        751 GTGGGATTCGGTGGCTCCTGCTTCCAGAAAGACATCCTCAACCTGCTG------------ 798

Qy        658 TGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGC 717

Db        799 TACATCTGCGAGTGCAATGGCCTCCCCGAGGTGGCCAACTACTGGAAGCAGGTTATCAAG 858

Qy        718 ATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACC 777

Db        859 ATCAACGACTACCAGAAGAGCCGGTTCGTGAACCGCGTGGTGTCGTCCATGTTCAACACC 918

Qy        778 GCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCC 837

Db        919 GTGGCCGGCAAGAAGATCGCCGTGCTGGGCTTCGCCTTCAAGAAGGACACCGGCGACACC 978

Qy        838 TACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTG 897

Db        979 AGGGAGACGCCCGCCATCGACGTCTGCAAGGGGCTGATCGGCGACAAGGCCAAGGTGAGC 1038

Qy        898 GACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC 957

Db       1039 ATCTACGACCCCCAGGTGACGGAGGACCAGGTCCAGCGCGACCTCGCCATGAGCAAGTTC 1098

Qy        958 CACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTAC 1017

Db       1099 GATTGGGATCACCCGGTCCATCTCCAGCCGATGAGCCCCACGGCCATCAAGCAGGTGTCC 1158

Qy       1018 TTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATC 1077

Db       1159 GTCGCCTGGGACGCCTACGAGGCGGCCAGGGACGCCCACGGCGTCTGCATCCTCACCGAG 1218

Qy       1078 AACAACGAGACCCCCGGCATCCGCTAC 1104

Db       1219 TGGGACGAGTTCAGGAGCCTCGACTAC 1245
```

RESULT 5
BM321451
LOCUS   BM321451   951 bp   mRNA   linear   EST 03-JAN-2002
DEFINITION rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321451
VERSION   BM321451.1 GI:18055857
KEYWORDS   EST.
SOURCE   Mastigamoeba balamuthi
  ORGANISM Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1 (bases 1 to 951)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE   The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT   Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 951   Std Error: 0.00
            POLYA=No.
FEATURES   Location/Qualifiers
     source   1..951
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match   3.5%; Score 85.6; DB 4; Length 951;
  Best Local Similarity   46.0%; Pred. No. 3.1e-05;
  Matches 323; Conservative   0; Mismatches 376; Indels   3; Gaps   1;

```
Qy       1728 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1787

Db        219 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 278

Qy       1788 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1847

Db        279 CATCGCGCAGGCCGGCGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 338

Qy       1848 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1907

Db        339 GGAGTGCACCTGGAAGGCCCTGTGCTTCGGCCCCTACCAGGGCCCTCAGATCATCGTCGA 398

Qy       1908 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1967

Db        399 CGACGGCGGTGACGCGACTCTGATGATCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 458

Qy       1968 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2027

Db        459 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 518

Qy       2028 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2087

Db        519 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 578

Qy       2088 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2147

Db        579 CGAGGAGACGACGACTGGCGTNATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 638

Qy       2148 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2207

Db        639 GTTCCCGGCCGTCAACGTCAACGAC---TCTNTCACCAAGAGCAAGTTCGACAACATCTA 695

Qy       2208 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2267

Db        696 CGGCTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGG 755

Qy       2268 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2327

Db        756 CAAGGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCG 815

Qy       2328 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387

Db        816 CGGCCAGGGCTGCCGCGTCATCGTGACGGAGATCGACCCCATCTGCGCGCTGCAGGCGTC 875

Qy       2388 CATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGA 2429

Db        876 GATGGCCGGCTTCGAGGTCAACACGCTCGAGGCGGGGCTCGA 917
```

RESULT 6
CL974989
LOCUS   CL974989   2031 bp   DNA   linear   GSS 21-SEP-2004
DEFINITION OsIFCC042895 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL974989
VERSION   CL974989.1 GI:52404497
KEYWORDS   GSS.
SOURCE   Oryza sativa (indica cultivar-group)

```
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2031)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..2031
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.5%;  Score 85.4;  DB 9;  Length 2031;
  Best Local Similarity   42.6%;  Pred. No. 3.6e-05;
  Matches  787;  Conservative    0;  Mismatches 1026;  Indels   36;  Gaps    5;

Qy        576 CGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCGTGAACATCATCGGCCG 635
Db        153 CGGCACGACGTACTCGTGCGTCGGAGTCTACCGGAACGGCCATGTCGAGATCATCGCCAA 212

Qy        636 CAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGT 695
Db        213 CGACCAGGGGAACCGGATCACGCCCTCGTGGGTCGCCTTCACCGACGGCGGCGAGCGGCT 272

Qy        696 GCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGA 755
Db        273 CATCGGTGAGGCCGCCAAGAACCAGGCGGCGGCCAACCCGGAGCGGACCATCTACGACGC 332

Qy        756 GGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCAC 815
Db        333 CAAGCGGCTCATCGGCCGGCAGTTCTCCGACGCCGAGGTGCAGCGCGACATGAAGCTGCT 392

Qy        816 CAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAG 875
Db        393 GCCGTTCGCCGTCGTCGACCGGAACGGCAAGCCGCACGTGCGCGTCGAGGTGAAGGACGG 452

Qy        876 CACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTG 935
Db        453 CGACGTGCG------------CGTGTTCAGCCCGGAGGAGGTGAGCGCCATGGTGCTCAC 500

Qy        936 GGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGT 995
Db        501 GCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCACCCGCGCCGTCGTCAC 560

Qy        996 GCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACAC 1055
Db        561 CGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGGACGCCGGCGTCATCGC 620

Qy       1056 CGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGT 1115
Db        621 CGGGCTCACCGTCGACCGCATCATCAACGAGCCCACCGCCGCCGCCATCGCCTACGGCAT 680

Qy       1116 GCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175
Db        681 CGACAAGAAGGGCGCCGAGAAGAACGTCCTCGTCTTCGACCTCGGCGGCGGCACGTTCGA 740

Qy       1176 GGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTG-- 1233
Db        741 CGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTGCCACCAACGGCGACAC 800

Qy       1234 ----GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCA 1289
Db        801 CCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACTTCGTCAAGGTCATCCG 860

Qy       1290 CCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCT 1349
Db        861 CCGGAAGCACGGCCGCGACATCACCGGCGACGCGCGCGCGCTGGGCAAGCTCCGCCGCGA 920

Qy       1350 GTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCT---GCC 1406
Db        921 GTGCGAGCGCGCCAAGCGCGCGCTCAGCAACCAGCACCAGGTGCGCGTCGAGGTCGAGTC 980

Qy       1407 CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGC 1466
Db        981 CCTGTTCGACGGCGTCGACTTGTCGGAGCCGCTCTCCCGGGCGCGGTTCGAGGAGCTCAA 1040

Qy       1467 CAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAA 1526
Db       1041 CAGCGACCTCTTCAAGAAGACGATGGTGCCGGTGAGGAAGGCCATGGCGGACGCCCGGCT 1100

Qy       1527 GGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCG 1586
Db       1101 GAGCAAGGGCGACATCGACGAGATCGTCCTCGTCGGCGGCAGCACCAGGATCCCCAAGGT 1160

Qy       1587 CGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGC 1646
Db       1161 GCAGCAGCTGCTCAAGGACTACTTCGGTGGCAAGGAGCCCAACCGCGGCGTCAACCCCGA 1220

Qy       1647 CGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAA 1706
Db       1221 CGAGGCCGTGGCGTACGGCGCCGCCGTGCAGGCCAGCATCATCAGCGGCCACGTCGACGA 1280

Qy       1707 GAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCA 1766
Db       1281 GAACACGGAGAGCATGATCCTCCTCGACGTCGCGCCGCTCACCCTCGGCCTGGAGACGGC 1340

Qy       1767 GCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCC 1826
Db       1341 CGGCGGCGTGATGGCCAAGCTGATCCCGCGCAACACGGTGGTGCCGACGAAGAAGACGCA 1400

Qy       1827 CAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCA 1886
Db       1401 GGTGTTCACCACGTACAAGGACAAGCAGACAACGGTGACCATCCAGGTGTTCGAGGGCGA 1460

Qy       1887 GGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTA 1946
Db       1461 GCGGAGCATGACGAGGGACAACCGGCTGCTCGGCAGGTTCGACCTCGCCGGCATCGCGCC 1520

Qy       1947 CCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAA 2006
Db       1521 GGCGCCGAGGGGGGCGCCGCAGATCGAGGTGACG---------TTCGAGGTGGACGCGAA 1571

Qy       2007 CCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGT 2066
Db       1572 CGGCATCCTCAGCGTGCTGGCCGCCGACAAGGCCACCGGCAGGTCGGAGAAGATCACCAT 1631

Qy       2067 GAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCA 2126
Db       1632 CTCCGGCGACGACCGCAAGATCAGCCAGGAGGAGATCGACCGGATGGTGCGCGAGGCGGA 1691

Qy       2127 GGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCA 2186
Db       1692 GGAGTTCGCCGAGGAGGACCGCCGGCACAGGGAGCAGGTGGACGCCCGGAACAGCCTGGA 1751

Qy       2187 GGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGAT------CATCGAGCAGCT 2240
Db       1752 GGCGTACGTCTACAACATCAAGAACACGCTCGGCGGCAAGATGGCGGACGCCATGGAGGG 1811

Qy       2241 GATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAA 2300
```

```
Db      1812 CGAGGAGAAGGACAAGGTGGAGGAGGCGGTGAGGGAGGCGTACGAGTGGCTGGACGGCAA 1871

Qy      2301 CGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCAT 2360

Db      1872 CCCGGACGCCGGCAAGGAGGAGTACGAGGAGAAGCTGAGGGAGCTGGAGGACGTGTGCAA 1931

Qy      2361 CGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGC 2409

Db      1932 CCCCGTCATGTCGGCGGTCTACCAGAGGTCCGGCGGCGGCGGCGGC 1980


RESULT 7
CL978463
LOCUS       CL978463                1725 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC031917 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL978463
VERSION     CL978463.1  GI:52411427
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1725)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1725
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.4%;  Score 84;  DB 9;  Length 1725;
  Best Local Similarity   43.7%;  Pred. No. 6.3e-05;
  Matches  470;  Conservative    0;  Mismatches  600;  Indels    6;  Gaps    2;

Qy       949 GGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGC 1008

Db        76 GTCCTCCGCACCCCGGACGCCCTCCACCGCATGTACGCGGAGGTGAGCCTCCTCGCCGAC 135

Qy      1009 GACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATC 1068

Db       136 CTCCGCCACGACGCCATCATCGCCTTCCACGCCTCCTGGGTCCACCCCTCCCGCCGCACC 195

Qy      1069 CCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGC 1128

Db       196 TTCAACTTCATCACCGAGCTCTTCTCCTCCGGCACCCTCCGCTCCTACCGCCTCCGCTAC 255

Qy      1129 TGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGC 1188

Db       256 CCGCGCGTCAGCCGCCGCGCCGTCGCCGCCTGGGCGCGCGCCATCCTCCACGGCCTCGCC 315

Qy      1189 GCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAG 1248

Db       316 TACCTCCACTCCCGCGGCGTCATCCACCGCGACCTCAAGTGCGACAACATCTTCGTCAAC 375

Qy      1249 ATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTC 1308

Db       376 GGCCACCTCGGCCAGGTCAAGATCGGCGACCTCGGCCTCGCCGCCGTCCTCCGCGGCTGC 435

Qy      1309 ACCACCCCCGACA---AGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365

Db       436 ACCTCCGCCCGCAGCGTCATCGGCACGCCGGAGTTCATGGCCCCCGAGATGTACGACGAG 495

Qy      1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425

Db       496 TGCTACGGCGTCGGCGTCGACGTCTACTCCTTCGGCATGTGCATGCTCGAGATGCTCACC 555

Qy      1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485

Db       556 AACGAGTACCCCTACAGCGAGTGCGACAACCCGGCGCAGATCTACAAGAAGGTCACCGCC 615

Qy      1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG-- 1543

Db       616 GGCAAGCTCCCCGACGCCTTCTACCTCCTCACCGACGCCGACGCCCGCCGCTTCATCGGC 675

Qy      1544 -TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAG 1602

Db       676 CGGTGCCTTGTCGACGCCGCCCACCGCCCCTCTGCCGAGGAGCTCCTGCTCGACCCCTTC 735

Qy      1603 CCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAG 1662

Db       736 CTCTCCCCGCCTCAGAACCACGACGATCACAATACCATTGCTCATGCCACTGCGCCGCCA 795

Qy      1663 GGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGC 1722

Db       796 CCGCCATTGCCATTGGCGTGCAGTAATTCCAGTGAGGAGCAAGAGGAGGAGGAGGCGCCG 855

Qy      1723 AAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTG 1782

Db       856 GCAGCAAAGACGACGGGCATGGCGATCACCGGGAAGCTGAACAAGGAGCACGACACCATC 915

Qy      1783 CAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCC 1842

Db       916 TTCCTCAAGGTGCAGATCGGCGGCGGCGGCAACGTGCGGAACATCTACTTCCCGTTCGAC 975

Qy      1843 ATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCC 1902

Db       976 GTGGCGAACGACACGGCGATGGAGGTGGCGACGGAGATGGTGAAGGAGCTGGACATCGCC 1035

Qy      1903 GAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAG 1962

Db      1036 GACCGGGAGCCCACGGAGATTGCCGCCATGATCGAGCAGGAGATCGTGAGGTTGGTACCA 1095

Qy      1963 CCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAA 2018

Db      1096 GGGTACAAGCAGCACGAGTATTCGTACGCGGATGACGACGACGATGATGACGTCAA 1151


RESULT 8
CL973991
LOCUS       CL973991                3069 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC025416 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL973991
VERSION     CL973991.1  GI:52402507
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 3069)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
```

```
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..3069
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match            3.3%;  Score 80.8;  DB 9;  Length 3069;
  Best Local Similarity  41.6%;  Pred. No. 0.00025;
  Matches  657;  Conservative    0;  Mismatches  912;  Indels    9;  Gaps    2;

Qy        37 ACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAG 96
Db       625 AGCGGCATCGGCAAGACGACGCTGGCGAGCTACCTCTACCGGAGGAACAGCATCCGGAAG 684

Qy        97 TGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAG 156
Db       685 CACTTCGCCTGCACGGCGTGGGTCCACGTCCCCCGCAAGTTCCGCTTCGCCGACCTCCTC 744

Qy       157 GGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCC 216
Db       745 CACGACGTCATCCGCCAGGCCAACGCCTCCTCCGGCGACGACGAGCTCGGCGGCGAGGC- 803

Qy       217 AACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAG 276
Db       804 --CGCCGACGACGACGTGCGGCGCCGGGGTGGATGGAGGCGGAGAGGCTCGTCAAGGCG 861

Qy       277 CAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGC 336
Db       862 CGCCTCGCCGCCGCGCTCCAGGGGAAGAGGTACCTCGTGGTGCTCGACGACGTCCGGAGC 921

Qy       337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAG 396
Db       922 AGGGAGGAGTGGGCCTTCTTCCTCGCCGCGCTCCCCGGCGGCCTCGCCGGCAGCTGCGTG 981

Qy       397 CGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGC 456
Db       982 CTCGTCACGACGCAGGTCAAGATCAGCGAGGAGTCGGAGCAACCCGTGGCCGCCGGCGGC 1041

Qy       457 GCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATC 516
Db      1042 GGCGTCGTGCGCACCAGGGAGCTCGGGAAGCTGCGGCCGGAGCAGGCCGCCGAGCTGTTC 1101

Qy       517 GGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGC 576
Db      1102 CGCCGGAGGGTGTACGGCCACGGCGAGCCGGAGCCGAAGAAGATGGAGCAGCTCAAGAGC 1161

Qy       577 GGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGC 636
Db      1162 CTCGTGGACTCCATGACGAAGGGCTCCAACCTGCCGCTCAACATCGTCATGCTCGCCGGA 1221

Qy       637 AACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTG 696
Db      1222 CTCTTGAGGTCCAAGAAGGAGGATGAATGGGAGGCCGTGATCCATAGCCTCGATGATACG 1281

Qy       697 CCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAG 756
Db      1282 CCACCCCCGACATCGCCGGAGCAGACGCCGCCGGAGGCGGAGCCGAAGAAGGACAGCCTC 1341

Qy       757 GAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACC 816
Db      1342 GATGATAGCGAGACGGCGCCGACATCGACGGAGCAAACGAAGAAGACGAAGAAGAGGAAG 1401

Qy       817 AAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGC 876
Db      1402 CAGATGGCGACATCGATGGACAAGATACTGACGGTGTGCATGGACGACCTGCCGACGCAT 1461

Qy       877 ACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936
Db      1462 CTGAAGCCGTGCTTCCTCTACTTCGCCGGGTTCACCGCGCAGACGCCGATCTGCGCCGGG 1521

Qy       937 GAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTG 996
Db      1522 AAGCTGGTCCGGCTGTGGGTCGCCGAGGGGTTCATGCAGGCGAAGAACGGGCAGACGGTG 1581

Qy       997 CTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACC 1056
Db      1582 GAGGAGCACGGCGAGGAGTGCCTCAAGGAGCTCATCTCCCGGTGCCTCGTCCAGCTCGTC 1641

Qy      1057 GCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTG 1116
Db      1642 GAGACGGACGCCGGCGGCGGCAGGGTCGCCGCCGTCAGCATCCACCAAGCGGTGCTCGAC 1701

Qy      1117 CTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTG 1176
Db      1702 TTCGTCCAGGCGGAGGCGCGCGACACCAACTTCCTCCACGTGCACAGCGGCGCGGCGGGC 1761

Qy      1177 GAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGC 1236
Db      1762 CTCTCCAACGGCGCCGCCCGCCGCCTCGCGCTCCGCAACACCTACGACACCGACCTCGCC 1821

Qy      1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
Db      1822 GTCATGCTCGAGGCGCCCAAGCTCCACACCCTCCTCTGCGACATCCCGGAGCGCG----- 1876

Qy      1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATG 1356
Db      1877 -CCGCCGCCGACGCCGCCGGCGCCGCGGCGTGGCAGAGAGCGCTCGAGCTCATCAATGGC 1935

Qy      1357 GGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1416
Db      1936 CGCGCCCCCACCTTCTCCGTCCACGGCTCGAGGTTCCTCCGCGTGATGGATCTCAAGGGC 1995

Qy      1417 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1476
Db      1996 GTGCGGCTGCCGCACCGGGAGTCGCTGCCGGAGGAGATCGGGTGGCTGATCCACCTCAGG 2055

Qy      1477 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1536
Db      2056 TACCTGGGCCTCTCGCACACCACCATGAGGCACCTCCCGAGCTCCGTCAAGAGGCTCCGG 2115

Qy      1537 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1596
Db      2116 AACCTGCAGACGCTGGACGTGTCGCGCACGGACGTCGAGGCGCTGCCGTGGCGGCTCTGG 2175

Qy      1597 CGCGAGCCCGTGCACGGC 1614
Db      2176 CGCAACCCCTCGCTCCGC 2193


RESULT 9
CK159167/c
LOCUS       CK159167                 869 bp    mRNA    linear   EST 05-DEC-2003
DEFINITION  FGAS040564 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA,
            mRNA sequence.
ACCESSION   CK159167
VERSION     CK159167.1  GI:38985053
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 869)
```

```
AUTHORS     Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
            Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
            Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
            Penniket,C., Roach,J.L. and Sarhan,F.
  TITLE     Functional Genomics of Abiotic Stress In Wheat and Canola Crops
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [128,636].
            Plate: TaLt537  row: N  column: 23.
FEATURES             Location/Qualifiers
     source          1. .869
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Wheat line PI 178383"
                     /db_xref="taxon:4565"
                     /lab_host="DH5 alpha"
                     /clone_lib="Triticum aestivum FGAS: TaLt5"
                     /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
                     subtractive hybridization) cDNA library from genotype
                     PI178383 cold hardened at 2 C for 21 days and 49 days
                     (equal amount of cDNA pooled together before subtraction,
                     tester) and subtracted against genotype Norstar cold
                     hardened at 2 C for 1 day (24 H)(driver). Modified Smart
                     cDNA (Clontech)priming and non-directional cloning"
ORIGIN

  Query Match             3.3%;  Score 80.6;  DB 7;  Length 869;
  Best Local Similarity   44.9%;  Pred. No. 0.00025;
  Matches  305;  Conservative    0;  Mismatches  374;  Indels    0;  Gaps    0;

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
Db        827 ATGGCGGCGGCCCGCACAACGAGGATCACCAACCAGGACGAGAACAACCACAACCACCAC 768

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
Db        767 CACAACGACGACAACAGCGACACAACGACCACCAACAACACCAACAACAACAACAACAAC 708

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
Db        707 ACCAACAACAACAACAACAACAACAACAACAACGACAACCACCACCACAACCACAACAAC 648

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
Db        647 AACAACAACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 588

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
Db        587 AACAACAACAACAACAACAACAACAACAACAACAACGACAACAACAACAACAACAACAAC 528

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
Db        527 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 468

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
Db        467 AACAACAACAACAACAACAACAACAACGACGACGACGACAACGACGACGACGACGACGAC 408

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
Db        407 GACAACAACGACAACAACGACAACAACAACAACGACAACAACAACAACAACAACAACAAC 348

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
Db        347 GACCACAACGACAACGACAACAACAACAACAACAACAACAACAACGACAACAACCACAAC 288

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
Db        287 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 228

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGC 1239
Db        227 AACAACAACAACAACAACGACGACAACGACGTCAACGATAACAACGACAACAACGACAAC 168

Qy       1240 GACCTGGAGATCGGCCAGC 1258
Db        167 AACGGCAACAACAACAAAC 149


RESULT 10
CL971508
LOCUS       CL971508                1941 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC021485 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL971508
VERSION     CL971508.1  GI:52397596
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1941)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .1941
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.3%;  Score 80.6;  DB 9;  Length 1941;
  Best Local Similarity   42.4%;  Pred. No. 0.00026;
  Matches  757;  Conservative    0;  Mismatches 1004;  Indels   24;  Gaps    5;

Qy        267 CCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGA 326
Db        165 CGCCGCCAAGAACCAGGTCGCCATGAACCCAACCAACACCGTCTTCGATGCGAAGCGGTT 224

Qy        327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCAC 386
Db        225 GATTGGTAGGAGGTTCTCCGACCCGTCCGTGCAGAGTGACATGAAGCTATGGCCGTTCAA 284

Qy        387 CCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCT 446
Db        285 GGTGGTGCCTGGCCCCGGCGACAAGCCGATGATCGTCGTCCAGTACAAGGGCGAGGAGAA 344

Qy        447 GGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCC 506
```

```
Db        345 GCAGTTCGCCGCGGAGGAGATCTCGTCCATGGTGCTGATCAAGATGAGGGAGATCGCGGA 404

Qy        507 CAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGAT 566

Db        405 GGCCTACCTCGGCTCGTCCATCAAGAACGCCGTGGTCACCGTCCCGGCCTACTTCAACGA 464

Qy        567 CGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACAT 626

Db        465 CTCGCAGAGGCAGGCCACCAAGGACGCCGGCGTCATCGCCGGGCTCAACGTGATGCGCAT 524

Qy        627 CATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCAT 686

Db        525 CATCAACGAGCCCACCGCCGCCGCCATCGCGTACGGCCTCGACAAGAAGGCGACGAGCAG 584

Qy        687 CGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCC 746

Db        585 CGGCGAGAAGAACGTGCTCATCTTTGACCTCGGCGGCGGCACGTTCGATGTGTCGCTGCT 644

Qy        747 CCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGG 806

Db        645 CACCATCGAGGAGGGCATCTTCGAGGTCAAGGCCACCGCCGGCGACACCCACCTCGGCGG 704

Qy        807 CAAGATCACCAAGATCGGCCCCGAGAACCCCTAC---AACACCCCCGTGTTCGCCATCAA 863

Db        705 CGAGGACTTCGACAACCGCATGGTGAACCACTTCGTGCAGGAGTTCAAGCGCAAGAACAA 764

Qy        864 GAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCAC 923

Db        765 GAAGGACATCAGCGGCAACCCGCGCGCGCTGCGGCGGCTCAGGACGGCGTGCGAGCGCGC 824

Qy        924 CCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAA 983

Db        825 CAAGCGCACGCTGTCGTCGACGGCGCAGACCACCATTGAGATCGACTCGCTGTACGAGGG 884

Qy        984 GAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTT 1043

Db        885 CATCGACTTCTACACGACGATCACCAGGGCGCGGTTCGAGGAGCTCAACATGGACCTGTT 944

Qy       1044 CCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTA 1103

Db        945 CCGCAAGTGCATGGAGCCCGTGGAGAAGTGCCTCCGCGACGCCAAGATGGACAAGAGCAG 1004

Qy       1104 CCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCAT 1163

Db       1005 CGTGCACGACGTCGTGCTCGTCGGCGGCTCCACCCGCATCCCCAAGGTGCAGCAGCTGCT 1064

Qy       1164 GACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCC 1223

Db       1065 CCAGGACTTCTTCAACGGCAAGGAGCTCTGCA------AGAGCATCAACCCCGACGAGGC 1118

Qy       1224 CCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCG 1283

Db       1119 CGTGGCGTACGGCGCCGCCGTCCAGGCCGCCATCCTGAGCGGCGAGGGCAACGAGAAGGT 1178

Qy       1284 CAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCC 1343

Db       1179 GCAGGACCTGCTCCTGCTCGACGTCACCCCGCTGTCCCTCGGCCTCGAGACCGCCGGCGG 1238

Qy       1344 CTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCA---GCCCATCGA 1400

Db       1239 CGTCATGACCGTGCTCATCCCGAGGAACACCACCATCCCGACCAAGAAGGAGCAGGTCTT 1298

Qy       1401 GCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAA 1460

Db       1299 CTCCACCTACTCCGACAACCAGCCCGGCGTCCTGATCCAGGTGTACGAGGGCGAGCGCGC 1358

Qy       1461 CTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA---------AGCT 1511

Db       1359 CAGGACCAAGGACAACAACCTCCTCGGCAAGTTCGAGCTCTCCGGCATCCCACCGGCGCC 1418

Qy       1512 GCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGA 1571

Db       1419 GCGCGGCGTGCCCCAGATCAACGTCTGCTTCGACATCGACGCCAACGGCATCCTGAACGT 1478

Qy       1572 GCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAG 1631

Db       1479 GTCCGCGGAGGACAAGACGACGGGGCAGAAGAACAAGATCACCATCACCAACGACAAGGG 1538

Qy       1632 CAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTA 1691

Db       1539 GCGGCTGAGCAAGGAGGAGATCGAGAAGATGGTGCAGGAGGCGGAGAAGTACAAGGCCGA 1598

Qy       1692 CCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACAC 1751

Db       1599 GGACGAGGAGCACAAGAAGAAGGTGGACGCCAAGAACGCCCTCGAGAACTACGCCTACAA 1658

Qy       1752 CAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGAT 1811

Db       1659 CATGCGCAACACCATCAAGGACGAGAAGATCGCCTCCAAGCTCGCCGCCGACGACAAGAA 1718

Qy       1812 CTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTG 1871

Db       1719 GAGGATCGAGGACGCCATCGATGGCGCCATCAGCTGGCTCGACACCAACCAGCTCGCCGA 1778

Qy       1872 GACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCT 1931

Db       1779 GGCCGACGAGTTCGAGGACA---AGATGAAGGAGCTGGAGGGCATCTGCAACCCGATCAT 1835

Qy       1932 GGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGT 1991

Db       1836 CGCCAAGATGTACCAGGGCCCCGGCGCCGACATGGCCGGAGGCATGGACGAGGACGCGCC 1895

Qy       1992 GGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTA 2036

Db       1896 GGCGGGCGGCAGCGGCGCAGGCCCCAAGATCGAGGAGGTCGACTA 1940
```

RESULT 11
BM320864

```
LOCUS       BM320864                1132 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM320864
VERSION     BM320864.1  GI:18055270
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1132)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1132   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1132
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.3%;  Score 80.2;  DB 4;  Length 1132;
```

```
  Best Local Similarity   45.3%;  Pred. No. 0.0003;
  Matches  375;  Conservative    0;  Mismatches  444;  Indels    9;  Gaps    2;

Qy        139 CGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGAC 198
Db         79 CGCCGCGAGGGCAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAG 138

Qy        199 TGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCC 258
Db        139 TACAACAGCCCCAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTC 192

Qy        259 CGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG 318
Db        193 TGCCAGATCGCCTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCAC 252

Qy        319 CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCC 378
Db        253 GAGCTCACCCGCTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACT 312

Qy        379 CAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAG 438
Db        313 GGCCTGCTGCTGGCCCGCCGTGTGCTGAAGAAGCTCAACCTCGACTCCAAGTACGAGGGT 372

Qy        439 GCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAG 498
Db        373 GTCAAGAAGGTCAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGT 432

Qy        499 TGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAG 558
Db        433 TTCAAGGCCCTGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCC 492

Qy        559 ATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCC 618
Db        493 GCCCTCAAGGGCATGTGCGACGGCGGCGTCAACGTCCCC---CACAGCGAGACCCGCTTC 549

Qy        619 GTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATC 678
Db        550 GTCGGCTTCAACGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTC 609

Qy        679 AGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAG 738
Db        610 GGCGGCCACGTCGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGAC 669

Qy        739 CAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAG 798
Db        670 CGCCAGTTCTCGCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTAC 729

Qy        799 AAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCC 858
Db        730 ACCGAGGCCCACAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAG 789

Qy        859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918
Db        790 CCCGAGGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAG 849

Qy        919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db        850 AACCGCGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCC 897

RESULT 12
BM320900
LOCUS       BM320900                1165 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM320900
VERSION     BM320900.1  GI:18055306
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1165)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1165   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..1165
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.2%;  Score 79.6;  DB 4;  Length 1165;
  Best Local Similarity   45.1%;  Pred. No. 0.00038;
  Matches  428;  Conservative    0;  Mismatches  506;  Indels   16;  Gaps    3;

Qy         90 CATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCG 149
Db         46 CGTCAAGAACAAGGCGTACTTCAAGCGCTTCCAGACCCAGTTCCGTCGCCGCCGCGAGGG 105

Qy        150 CAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCG 209
Db        106 CAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAGTACAACAGCCC 165

Qy        210 CCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCC 269
Db        166 CAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTCTGCCAGATCGC 219

Qy        270 CAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAA 329
Db        220 CTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCACGAGCTCACCCG 279

Qy        330 CCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCT 389
Db        280 CTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACTGGCCTGCTGCT 339

Qy        390 GTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGA 449
Db        340 GGCCCGCCGTGTGCTGAAGAAGCTCAACCTCTACTCCAAGTACGAGGGTGTCAAGAAGGT 399

Qy        450 CACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAA 509
Db        400 CAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGTTCAAGGCCC-- 457

Qy        510 GATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGA 569
Db        458 --TGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCCGCCCTCAA 515

Qy        570 GATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCAT 629
Db        516 GGGCATGTGCGACGGCGGCGTCAACGTCCCCCACAGCGAGACCCGCTTCGTCGGCTTCAA 575

Qy        630 CGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGA 689
Db        576 CGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTCGGCGGCCACGT 635

Qy        690 GACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCT 749
Db        636 CGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGACCGCCAGTTCTC 695
```

```
Qy       750 GACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAA 809
Db       696 GCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTACACCGAGGCCCA 755
Qy       810 GATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAA 869
Db       756 ------CAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAGCCCGA 809
Qy       870 GGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGA 929
Db       810 GGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAGAACCG 869
Qy       930 CTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGT 989
Db       870 CGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCCAAGAGCCAGTAAATTCC 929
Qy       990 GACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGG 1039
Db       930 CGGTGCCCTGTACACCGCCATCTCCCGCTGCGGCGTTGCGCTGCTNGCGG 979
```

RESULT 13
CL962901
LOCUS CL962901 11691 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFCC008219 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL962901
VERSION CL962901.1 GI:52380530
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 11691)
AUTHORS Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .11691
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Query Match 3.2%; Score 79.6; DB 9; Length 11691;
Best Local Similarity 43.5%; Pred. No. 0.00044;
Matches 587; Conservative 0; Mismatches 739; Indels 24; Gaps 4;

```
Qy       150 CAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCG 209
Db      2991 CAAGAAGGCCTTCGACAAGCAGGGCCTGAAGGTCGAGCTGGGCGTGAAGATCGGCGAAGT 3050
Qy       210 CCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCC 269
Db      3051 GAAGACGGGCGGCGACGGCGTGACCGTGGCCTACACCGATGCCAAGGGCGAGGCCCGCGA 3110
Qy       270 CAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAA 329
Db      3111 ACTGGCGGTGGACAAGCTCATCGTCTCCATCGGCCGCGTGCCCAACACCATCGGCCTCAA 3170
Qy       330 CCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCT 389
Db      3171 CCCCGAAGCCGTGGGCCTGAAGCTCGACGAGCGCGGCGCCATCGTGGTCGATGCCGACTG 3230
Qy       390 GTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGG- 448
Db      3231 CAAGACCAACCTGCCCGGCGTGTGGGCGGTGGGCGACGTGGTGCGCGGCCCGATGCTGGC 3290
Qy       449 --ACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCC 506
Db      3291 GCACAAGGCCGAGGAAGAGGGCGTGGCGGTGGCCGAGCGCATCGCGGGCCAGCATGGCCA 3350
Qy       507 CAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGAT 566
Db      3351 CGTGAACTTCAACACCATCCCCTGGGTCATCTACACCAGCCCCGAGATCGCCTGGGTGGG 3410
Qy       567 CGAGATCTGCGGCAAG---AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAA 623
Db      3411 CCGCACCGAGCAGCAGCTCAAGGCCGACGGCGTGCAGTACAAGGCCGGCACCTTCCCGTT 3470
Qy       624 CATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCC 683
Db      3471 CCTGGCGAACGGCCGCGCCCGTGCGCTGGGCGACACGACCGGCATGGTCAAGTTCCTGGC 3530
Qy       684 CATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTG 743
Db      3531 CGATGCGGCCACGGACGAGATCCTGGGCGTGCACATCGTGGGCCCGATGGCCAGCGAGCT 3590
Qy       744 GCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGA 803
Db      3591 GATCTCCGAGGCCGTGGTGGCCATGGAGTTCAAGGCCAGCGCCGAGGACATCGCGCGCAT 3650
Qy       804 GGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAA 863
Db      3651 CTGCCACGCCCATCCCTCGCTTTCCGAGGCGACCAAGGAAGCGGCCCTGGCCGCGTATGA 3710
Qy       864 GAAGAAGGACAGCACCAAGTGGCGCA-----------AGCTGGTGGACTTCCGCGAGCT 911
Db      3711 GGCCGAGCTGGCCGCCAAGGGGTTCCGCAGCGACCCGGCCCAGCTGCGCGCCGTCGAGGC 3770
Qy       912 GAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCT 971
Db      3771 GCTGCAGCGCTGCGCGGACGACTGGGCCGCCTACAAGGAAAAGCGCTCCAACGCGCTCAA 3830
Qy       972 GAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCT 1031
Db      3831 GAAGCTCATCCACCGCCCCGAGGTGCCGCGCGGCGTGTACATGTACGGGGGCGTGGGCCG 3890
Qy      1032 GGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCC 1091
Db      3891 TGGCAAGAGCTTCCTGATGGACTGCTTCTTCAACGCCGTGCCGATCCGCCGCAAGGTGCG 3950
Qy      1092 CGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTT 1151
Db      3951 GCTGCATTTCCACGAATTCATGCGTGAGGTGCACCGCGAACTGGCGGGCCTGCAGGGCAC 4010
Qy      1152 CCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGAT 1211
Db      4011 CGTGAACCCGCTGGACGTGCTGGGCGAGCGCATCGCCAAGCGCTACAAGCTCATCTGCTT 4070
Qy      1212 CTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGAT 1271
Db      4071 CGACGAATTCCACGTGGCCGACATCACGGACGCCATGATCCTGCACCGCC------TGCT 4124
Qy      1272 CGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCA 1331
Db      4125 GGCGGCGCTGTTCGACAACGGCGTGGGCTTCGTCACCACCTCGAACTTCAAGCCCGACGA 4184
Qy      1332 GAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCA 1391
Db      4185 CCTGTACCCGGGCGGCCTGCACCGCGACCGCATCCTGCCGGCCATCGATCTGCTCAACGA 4244
```

```
Qy       1392 GCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGG 1451

Db       4245 GAAGCTGGAGGTGGTGAACGTGGACAACGGCACCGACTACCGGCGCCGCACGCTGGAGCA 4304

Qy       1452 CAAGCTGAACTGGGCCAGCCAGATCTACCC 1481

Db       4305 GGTCCGGCTCTACCACTGCCCGCTGGGCCC 4334


RESULT 14
BM321430
LOCUS       BM321430                867 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
            sequence.
ACCESSION   BM321430
VERSION     BM321430.1  GI:18055836
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 867   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..867
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.2%;  Score 79;  DB 4;  Length 867;
  Best Local Similarity   45.7%;  Pred. No. 0.00048;
  Matches  314;  Conservative    0;  Mismatches  370;  Indels    3;  Gaps    1;

Qy         58 CAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAG 117

Db        131 CCGCACAAGATGCGCGAGTGCCTGCCGGTCATCATCCTGCTGCGCAACAGGTTGAAGTAC 190

Qy        118 GGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAG 177

Db        191 GCGCTGACCCGCCGTGAGGTCACCTCGATCGTGATGCAGCGCCTGATCAAGATCGACGGC 250

Qy        178 GAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTG 237

Db        251 AAGGTCCGCACCGACACGCACTTCCCCGCGGGCTTCATGGACGTCGTCTCGATCGACAAG 310

Qy        238 GCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCC 297

Db        311 ACCGACGAGCACTTCCGCCTGCTCTACGACACCAAGGGCCGCTTCCAGGCGCACCGCATC 370

Qy        298 ACCAGCCGCGAGCTGCAGGT---GCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAG 354

Db        371 AACTCGACCGAGGCCAAGTTCAAGCTCGGCAAGGTCCGCCGCGTGCAGCTCGGCAACAAG 430

Qy        355 CGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATC 414

Db        431 GGCATCCCGTACCTGGTGACCCACGACGCGCGCACGATCCGCTACCCCAACCCCGACATC 490

Qy        415 AAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTG 474

Db        491 AAGGTCAACGACACGGTCAAGATCGACCTGCGCTCGGGCAAGATCATCGACTTCGTCAGG 550

Qy        475 GAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTC 534

Db        551 TTCGAGATCGGCAACCTCGTCATGATCACTGGCGGACGCAACCTTGGCCGCGTCGGCGTC 610

Qy        535 ATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGC 594

Db        611 ATTGTGCGCCGCGAGAAGCACGAGGGCTCGTTCGAGATCATCCACGTCAAGGACGCCGTC 670

Qy        595 ACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654

Db        671 GGCCACCAGTTCGCGACGCGCCTGACCAACGTCTTCGTGATCGGCAAGGGCACCAAGTCC 730

Qy        655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714

Db        731 CTCGTCACGCTGCCCGCCGGCAAGGGCATCAAGAAGTCGATCATCGAGGAGTTCCAGGCG 790

Qy        715 GGCATGGACGGCCCCAAGGTGAAGCAG 741

Db        791 CGCCACGCGCACAGGACCGAGCAGGAG 817


RESULT 15
CL969033
LOCUS       CL969033               2682 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC017680 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL969033
VERSION     CL969033.1  GI:52392695
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2682)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..2682
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.2%;  Score 78.6;  DB 9;  Length 2682;
  Best Local Similarity   42.1%;  Pred. No. 0.00061;
  Matches  453;  Conservative    0;  Mismatches  624;  Indels    0;  Gaps    0;

Qy        324 CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGAT 383

Db        501 CTACATCGAGAGCATCGTCGTCGAGGGGTTCCCGACCGGCCCGGCGCACTTCACCTGCAA 560
```

```
Qy        384 CACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCT 443
Db        561 CTCGTGGGTGCAGCCCACGCGCGTGAGCCGCGACCGGCGCGTGTTCTTCAGCAACCGGCC 620

Qy        444 GCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAA 503
Db        621 GTACCTGCCGTCGGAGACGCCGCCGGGGCTGAGGGAGCTCCGGCTCCGGGAGCTCGCCGA 680

Qy        504 GCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCT 563
Db        681 CCTCCGCGGCGACGGCACCGGCGAGCGGAGGATCACCGACCGGGTGTACGACTACGACGT 740

Qy        564 GATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAA 623
Db        741 GTACAACGACCTGGGCAACCCGGACAAGGGCGTCGCCTCCGCCCGCCCCGTCCTCGGCGG 800

Qy        624 CATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCC 683
Db        801 CGAGCAGATGCCGTACCCGCGGCGGATGCGCACCGGCCACGCGAGCGCGGAGAGCAGGGT 860

Qy        684 CATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTG 743
Db        861 GGAGTACCCGGAGCCCATCTACGTGTCGCGGGACGAGGAGTTCGAGGAGGGCAAGAACGA 920

Qy        744 GCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGA 803
Db        921 GATGCTGTCGGAGGGCGCGATCAAGGCGCTGCTCCACAACTTCATGCCGCTGCTCGTCTC 980

Qy        804 GGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAA 863
Db        981 CTCCGTCTCGCCGGACATCCGCGACTTCGCCGGCTTCCACGACGTCGACAACCTCTTCAA 1040

Qy        864 GAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCAC 923
Db       1041 GGAGGGCCTCCGCCTCAAGCAGGCGCTCCACGACCAGCTCTTCCAGAAGATCCCCTTCGT 1100

Qy        924 CCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAA 983
Db       1101 CCGCAAGATCCAGGAGAACAGCGAGGGCCTCCTCCGCTACGACACCCCCGACATCATCAA 1160

Qy        984 GAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTT 1043
Db       1161 GAAGGACAAGTTCGCGTGGCTGCGTGACGACGAGTTCGCGAGGCAGGCGCTTGCCGGCAT 1220

Qy       1044 CCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTA 1103
Db       1221 CAACCCCGTCAACATTGAGCGTCTCCAGGCGTTCCCGCCGGTGAGCAAGCTCGACCCGGC 1280

Qy       1104 CCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCAT 1163
Db       1281 CGTGTACGGCCCGCCGGAGTCGGCGATCACGGAGGAGCACATCATCGGGCACCTCGACGG 1340

Qy       1164 GACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCC 1223
Db       1341 CATGTCGGTGCAGGAGGCGGTGGAGGGGGAGCAGGCTGTACATGCTGGATTACCACGACAT 1400

Qy       1224 CCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCG 1283
Db       1401 CTTCCTCCCGTTCCTGGACAGGATCAACGCCCAGGACGGCCGGAAGGCGTACGGCACGCG 1460

Qy       1284 CAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCC 1343
Db       1461 CGCGGTCTTCTTCCTGACGGCCGCCGGCACGCTGAAGCCGATCGCCATCGAGCTGTGCCT 1520

Qy       1344 CTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGA 1400
Db       1521 GCCGCCGATGACAGACGGCTGCAAGCGCGCCAAGCGGGTGTTCACGCCGCCGGCCGA 1577
```

Search completed: June 2, 2005, 02:53:26
Job time : 7863 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         June  1, 2005, 07:40:05 ; Search time 7875.14 Seconds
                                           (without alignments)
                                           11933.846 Million cell updates/sec

Title:          US-09-610-313B-30
Perfect score:  2469
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2469

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters:      68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_htc:*
                4:  gb_est3:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_gss1:*
                9:  gb_gss2:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	136.4	5.5	330	9	CL293849	CL293849 02S0349-0
	2	97.8	4.0	2031	9	CL974989	CL974989 OsIFCC042
	3	97.2	3.9	2886	9	CL967755	CL967755 OsIFCC015
	4	91.4	3.7	1509	9	CL959255	CL959255 OsIFCC002
	5	90.8	3.7	2598	3	AY103647	AY103647 Zea mays
	6	86.6	3.5	1398	9	CL961989	CL961989 OsIFCC006
	7	85.6	3.5	951	4	BM321451	BM321451 rockefell
c	8	85	3.4	869	7	CK159167	CK159167 FGAS04056
	9	84	3.4	3069	9	CL973991	CL973991 OsIFCC025
	10	83.8	3.4	1941	9	CL971508	CL971508 OsIFCC021
	11	80.2	3.2	1132	4	BM320864	BM320864 rockefell
	12	79.6	3.2	1165	4	BM320900	BM320900 rockefell
	13	79	3.2	867	4	BM321430	BM321430 rockefell
	14	78.4	3.2	11691	9	CL962901	CL962901 OsIFCC008
	15	78.2	3.2	1962	9	CL961326	CL961326 OsIFCC005
	16	78.2	3.2	2853	9	CL974397	CL974397 OsIFCC025
	17	78	3.2	1485	9	CL970981	CL970981 OsIFCC020
	18	76.6	3.1	1550	4	BM321022	BM321022 rockefell
	19	76.6	3.1	2559	9	CL982027	CL982027 OsIFSC046
	20	76.2	3.1	1401	9	CL962721	CL962721 OsIFCC038
	21	75.6	3.1	671	6	CA093222	CA093222 SCCCCL300
	22	75.4	3.1	2299	3	AY106831	AY106831 Zea mays
	23	75.2	3.0	854	7	CK777127	CK777127 964131 MA
	24	75	3.0	862	4	BM321023	BM321023 rockefell
	25	75	3.0	3822	9	CL972913	CL972913 OsIFCC023
	26	74.6	3.0	914	8	BZ568300	BZ568300 pacs2-164
	27	74.6	3.0	2010	9	CL962831	CL962831 OsIFCC008
	28	74.6	3.0	2313	9	CL982362	CL982362 OsIFSC047
	29	74.6	3.0	2682	9	CL969033	CL969033 OsIFCC017
	30	74.4	3.0	3134	3	AY109500	AY109500 Zea mays
	31	74.2	3.0	1680	9	CL982770	CL982770 OsIFSC049
	32	74	3.0	1290	9	CL972679	CL972679 OsIFCC023
	33	74	3.0	2072	3	CR603312	CR603312 full-leng
c	34	73.8	3.0	889	7	CK159613	CK159613 FGAS04107
c	35	73.6	3.0	757	9	CC678788	CC678788 OGWII71TV
	36	73.6	3.0	2028	9	CL979437	CL979437 OsIFCC033
	37	73.2	3.0	853	4	BM321393	BM321393 rockefell
	38	73	3.0	1689	9	CL972373	CL972373 OsIFCC022
c	39	72.8	2.9	566	4	BM587428	BM587428 170006873
	40	72.8	2.9	1575	9	CL979927	CL979927 OsIFCC045
	41	72.6	2.9	788	6	CB643171	CB643171 OSJNEb03L
	42	72.4	2.9	753	9	CC675888	CC675888 OGWCO51TH
c	43	72.4	2.9	801	4	BJ375394	BJ375394 BJ375394
	44	72.2	2.9	688	6	CB648640	CB648640 OSJNEb12C
	45	72.2	2.9	764	6	CB651670	CB651670 OSJNEb16L

ALIGNMENTS

```
RESULT 1
CL293849/c
LOCUS       CL293849                330 bp    DNA     linear   GSS 12-FEB-2004
DEFINITION  02S0349-08A1-C03 UniformMu MuTAIL Library Zea mays genomic clone
            02S0349-08A1-C03, genomic survey sequence.
ACCESSION   CL293849
VERSION     CL293849.1  GI:42541978
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Latshaw,S., Tan,B-.C., Settles,A.M. and McCarty,D.R.
  TITLE     Sequence tagged transposon insertions from the UniformMu maize
            population
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Donald R. McCarty
            Plant Molecular and Cellular Biology Program
            University of Florida
            PO 110690 Gainesville, FL 32611-0690, USA
            Tel: 352-392-1928 x322
            Email: drm@ufl.edu
            Sequence flanking probable Mu insertion site in UniformMu line:
            02S0349-08, Primer set: A
            Class: transposon insertion site.
FEATURES             Location/Qualifiers
     source          1..330
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="W22 (ACR, bz1-m9)"
                     /cultivar="UniformMu"
                     /db_xref="taxon:4577"
                     /clone="02S0349-08A1-C03"
                     /clone_lib="UniformMu MuTAIL Library"
                     /note="Vector: TOPO-PCR4; DNA flanking Mu transposon
                     insertions in Mu inactive lines were extracted from the
                     UniformMu maize population by the thermo assymmetric
                     interlaced PCR (TAIL) protocol using primers specific for
                     the Mu terminal inverted repeat and a set of 16 arbitrary
                     primers. Amplicons were size enriched using Sepharose 400
                     spin columns and cloned into the TOPO PCR4 vector."
ORIGIN

  Query Match             5.5%;  Score 136.4;  DB 9;  Length 330;
  Best Local Similarity   66.4%;  Pred. No. 2.3e-14;
```

```
  Matches  211;  Conservative    0;  Mismatches  106;  Indels    1;  Gaps    1;

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db        329 GGGGGGCAGCTGGAAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAA 270

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db        269 ATGACTTTGACAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAA 210

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db        209 GTAAGACAGTATGATCAGGTACCCATAGAAATCTGTGGGCATAAAGCTATAGGTACGGTA 150

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db        149 TTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGG-TGC 91

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db         90 ACCTTAAATTTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATG 31

Qy        721 GACGGCCCCAAGGTGAAG 738
Db         30 GATGGCCCAAAAGTAAGG 13
```

RESULT 2
CL974989

```
LOCUS       CL974989               2031 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC042895 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL974989
VERSION     CL974989.1  GI:52404497
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2031)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..2031
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             4.0%;  Score 97.8;  DB 9;  Length 2031;
  Best Local Similarity   42.6%;  Pred. No. 2.2e-07;
  Matches  787;  Conservative    0;  Mismatches 1032;  Indels   30;  Gaps    4;

Qy        576 CGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCG 635
Db        153 CGGCACGACGTACTCGTGCGTCGGAGTCTACCGGAACGGCCATGTCGAGATCATCGCCAA 212

Qy        636 CAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGT 695
Db        213 CGACCAGGGGAACCGGATCACGCCCTCGTGGGTCGCCTTCACCGACGGCGGCGAGCGGCT 272

Qy        696 GCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGA 755
Db        273 CATCGGTGAGGCCGCCAAGAACCAGGCGGCGGCCAACCCGGAGCGGACCATCTACGACGC 332

Qy        756 GGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCAC 815
Db        333 CAAGCGGCTCATCGGCCGGCAGTTCTCCGACGCCGAGGTGCAGCGCGACATGAAGCTGCT 392

Qy        816 CAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAG 875
Db        393 GCCGTTCGCCGTCGTCGACCGGAACGGCAAGCCGCACGTGCGCGTCGAGGTGAAGGACGG 452

Qy        876 CACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTG 935
Db        453 CGACGTGCG-----------CGTGTTCAGCCCGGAGGAGGTGAGCGCCATGGTGCTCAC 500

Qy        936 GGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGT 995
Db        501 GCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCACCCGCGCCGTCGTCAC 560

Qy        996 GCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACAC 1055
Db        561 CGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGGACGCCGGCGTCATCGC 620

Qy       1056 CGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGT 1115
Db        621 CGGGCTCACCGTCGACCGCATCATCAACGAGCCCACCGCCGCCGCCATCGCCTACGGCAT 680

Qy       1116 GCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175
Db        681 CGACAAGAAGGGCGCCGAGAAGAACGTCCTCGTCTTCGACCTCGGCGGCGGCACGTTCGA 740

Qy       1176 GGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTA 1235
Db        741 CGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTGCCACCAACGGCGACAC 800

Qy       1236 CGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCA 1295
Db        801 CCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACTTCGTCAAGGTCATCCG 860

Qy       1296 CCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCT 1355
Db        861 CCGGAAGCACGGCCGCGACATCACCGGCGACGCGCGCGCGCTGGGCAAGCTCCGCCGCGA 920

Qy       1356 GTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCT---GCC 1412
Db        921 GTGCGAGCGCGCCAAGCGCGCGCTCAGCAACCAGCACCAGGTGCGCGTCGAGGTCGAGTC 980

Qy       1413 CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGC 1472
Db        981 CCTGTTCGACGGCGTCGACTTGTCGGAGCCGCTCTCCCGGGCGCGGTTCGAGGAGCTCAA 1040

Qy       1473 CAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAA 1532
Db       1041 CAGCGACCTCTTCAAGAAGACGATGGTGCCGGTGAGGAAGGCCATGGCGGACGCCCGGCT 1100

Qy       1533 GGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCG 1592
Db       1101 GAGCAAGGGCGACATCGACGAGATCGTCCTCGTCGGCGGCAGCACCAGGATCCCCAAGGT 1160

Qy       1593 CGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGC 1652
Db       1161 GCAGCAGCTGCTCAAGGACTACTTCGGTGGCAAGGAGCCCAACCGCGGCGTCAACCCCGA 1220

Qy       1653 CGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAA 1712
Db       1221 CGAGGCCGTGGCGTACGGCGCCGCCGTGCAGGCCAGCATCATCAGCGGCCACGTCGACGA 1280

Qy       1713 GAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCA 1772
```

```
Db     1281 GAACACGGAGAGCATGATCCTCCTCGACGTCGCGCCGCTCACCCTCGGCCTGGAGACGGC 1340

Qy     1773 GCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCC 1832
Db     1341 CGGCGGCGTGATGGCCAAGCTGATCCCGCGCAACACGGTGGTGCCGACGAAGAAGACGCA 1400

Qy     1833 CAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCA 1892
Db     1401 GGTGTTCACCACGTACAAGGACAAGCAGACAACGGTGACCATCCAGGTGTTCGAGGGCGA 1460

Qy     1893 GGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTA 1952
Db     1461 GCGGAGCATGACGAGGGACAACCGGCTGCTCGGCAGGTTCGACCTCGCCGGCATCGCGCC 1520

Qy     1953 CCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAA 2012
Db     1521 GGCGCCGAGGGGGGCGCCGCAGATCGAGGTGACG---------TTCGAGGTGGACGCGAA 1571

Qy     2013 CCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGT 2072
Db     1572 CGGCATCCTCAGCGTGCTGGCCGCCGACAAGGCCACCGGCAGGTCGGAGAAGATCACCAT 1631

Qy     2073 GAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCA 2132
Db     1632 CTCCGGCGACGACCGCAAGATCAGCCAGGAGGAGATCGACCGGATGGTGCGCGAGGCGGA 1691

Qy     2133 GGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCA 2192
Db     1692 GGAGTTCGCCGAGGAGGACCGCCGGCACAGGGAGCAGGTGGACGCCCGGAACAGCCTGGA 1751

Qy     2193 GGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGAT------CATCGAGCAGCT 2246
Db     1752 GGCGTACGTCTACAACATCAAGAACACGCTCGGCGGCAAGATGGCGGACGCCATGGAGGG 1811

Qy     2247 GATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAA 2306
Db     1812 CGAGGAGAAGGACAAGGTGGAGGAGGCGGTGAGGGAGGCGTACGAGTGGCTGGACGGCAA 1871

Qy     2307 CGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCAT 2366
Db     1872 CCCGGACGCCGGCAAGGAGGAGTACGAGGAGAAGCTGAGGGAGCTGGAGGACGTGTGCAA 1931

Qy     2367 CGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGC 2415
Db     1932 CCCCGTCATGTCGGCGGTCTACCAGAGGTCCGGCGGCGGCGGCGGCGGC 1980
```

```
RESULT 3
CL967755
LOCUS       CL967755                2886 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC015718 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL967755
VERSION     CL967755.1  GI:52390149
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2886)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..2886
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.9%;  Score 97.2;  DB 9;  Length 2886;
  Best Local Similarity   43.2%;  Pred. No. 2.8e-07;
  Matches  712;  Conservative    0;  Mismatches  903;  Indels   33;  Gaps    4;

Qy      189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
Db      219 GGTCATGGAGGCCGCCGCCATCATGGCCATCGTCCTCGCCAACGGCGGCGGCAGGCCGCC 278

Qy      249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
Db      279 CGACTGGCAGGACTTCGTCGGCATCGTCGTCCTCCTCGTCATCAACTCCACCATCAGTTT 338

Qy      309 GCTGCAGGTGCGC---GGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCAC 365
Db      339 CATCGAGGAGAACAATGCCGGGAACGCCGCCGCGGCGCTCATGGCCGGCCTCGCGCCCAA 398

Qy      366 CCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGG 425
Db      399 GACCAAGGTGCTCAGGGACGGGAAATGGCAGGAGCAGGACGCGTCCATCCTCGTTCCCGG 458

Qy      426 CCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAG 485
Db      459 CGACATCATCAGCATCAAGCTCGGCGACATCATCCCCGCCGACGCGCGGCTGCTCGAGGG 518

Qy      486 CCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCG 545
Db      519 CGACCCGCTCAAGGTCGACCAGGCGGCGCTCACCGGCGAGTCGATGCCCGTCAACAAGCA 578

Qy      546 CCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGAT 605
Db      579 CGCCGGGCAGGGCGTCTTCTCCGGGTCCACCGTCAAGCAGGGCGAGATCGAGGCCGTCGT 638

Qy      606 CGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCT 665
Db      639 CATCGCCACCGGCGTGCACACCTTCTTCGGCAAGGCGGCGCACCTGGTGGACAGCACCAA 698

Qy      666 GAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGG 725
Db      699 CAACATCGGCCACTTCCAGCTCGTGCTCACGGCCATCGGCAACTTCTGCATCATCTCCAT 758

Qy      726 CCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTG 785
Db      759 CGGCGTCGGCATGATAATCGAGATCATCGTTATGTACCCGATCCAGCACCGCGCGTACCG 818

Qy      786 CGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACAC 845
Db      819 CGACGGCATCGACAACCTGCTCGTCCTGCTCATCGGCGGCATCCCCATCGCCATGCCCAC 878

Qy      846 CCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCG 905
Db      879 CGTGCTGTCCGTCACCATGGCCATCGGCTCCCACCGGCTGTCGCAGCAGGGCGCCATCAC 938

Qy      906 CGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGC 965
Db      939 CAAGCGGATGACCGCCATCGAGGAGATGGCCGGCATGGACGTGCTCTGCAGCGACAAGAC 998

Qy      966 CGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGT 1025
Db      999 CGGGACGCTGACGCTCAACAAGCTCACCGTCGACAAGACGCTGATCGAGGTGTACGGGCG 1058
```

```
Qy       1026 GCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGA 1085
Db       1059 GGGGCTGGACAAGGACTCGGTGCTCCTGTACGCGGCGAGGGCGTCCCGCGTCGAGAACCA 1118

Qy       1086 GACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAG 1145
Db       1119 GGACGCCATTGACACATGCATCGTGGGCATGCTCGCCGACCCCAAGGAGGCCCGCGCCGG 1178

Qy       1146 CATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGAT 1205
Db       1179 CAT---------CAAGGAGGTCCACTTCCTCCCGTTCAACCCCGTCGAGAAGCGCACGGC 1229

Qy       1206 CGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCA 1265
Db       1230 CATCACCTACATCGACGGCAATGGCGAATGGCACAGGATCAGCAAGGGCGCGCCGGAGCA 1289

Qy       1266 CCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGA 1325
Db       1290 GATCATCGAGCTGTGCAAGATGAGCAAGGACGCCGAGAAGAAGGTGCACACGCTGATCGA 1349

Qy       1326 CAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAA 1385
Db       1350 CCAGTACGCGGACCGCGGCCTCCGGTCGCTGGGCGTGTCGTACCAGAAGGTGCCGGAGAA 1409

Qy       1386 GTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCA 1445
Db       1410 G--------------AGCAAGGAGAGCGAGGGCGAGCCATGGCAGTTCGTCGGCCTCCT 1454

Qy       1446 GAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCA 1505
Db       1455 GCCGCTGTTCGACCCGCCGAGGCACGACAGCGCGGAGACCATCCGCCGCGCGCTGCACCT 1514

Qy       1506 GCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGA 1565
Db       1515 CGGCGTGAACGTGAAGATGATCACCGGCGACCAGCTCGCCATCG------GGAAGGAGAC 1568

Qy       1566 GGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTA 1625
Db       1569 GGCGCGGCGGCTCGGGATGGGCACCAACATGTACCCGTCGACGACGCTGCTGGGCGACAA 1628

Qy       1626 CTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGAC 1685
Db       1629 GAGCAGCGAGATGAGCGGCCTCCCAATCGACGAGCTGATCGAGAAGGCCGACGGCTTCGC 1688

Qy       1686 CTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCG 1745
Db       1689 CGGCGTGTTCCCGGAGCACAAGTACGAGATCGTGAAGCGCCTCCAGGACCGGAAGCACAT 1748

Qy       1746 CACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGA 1805
Db       1749 CTGCGGCATGACCGGCGACGGCGTGAACGACGCGCCGGCGCTGAAGAAGGCCGACATCGG 1808

Qy       1806 GAGCATCGTGATCTGGGGCAAGACCCCC 1833
Db       1809 CATCGCCGTCGACGACGCCACGGACGCC 1836
```

```
RESULT 4
CL959255
LOCUS       CL959255                1509 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC002520 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL959255
VERSION     CL959255.1  GI:52373273
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1509)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1509
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.7%;  Score 91.4;  DB 9;  Length 1509;
  Best Local Similarity   43.8%;  Pred. No. 3e-06;
  Matches  395;  Conservative    0;  Mismatches  506;  Indels    0;  Gaps    0;
```

```
Qy        430 ATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTG 489
Db          1 ATGGACGACATCATGCGGGCCATCTTCGCCTGCATCCCCGGCATGCGGTCACCCGCGTCG 60

Qy        490 CCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAG 549
Db         61 CCCGGCGGCTGCCTCTCCGCTGCCTTTGGCGGCGGGGACGACGACGGCGAGGAGGGGGAG 120

Qy        550 TACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGC 609
Db        121 GGGGAGGGGGAGGACCGCATCAGCCGCCTCCCGGACGCGGTGCTGTCGAGCATCGTCTCC 180

Qy        610 CCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAAC 669
Db        181 CGCCTCCCCGTCAAGGACGGCGCCCGCACCGCCGTGCTGTCCCCGCGCTGGCGCCGCGTC 240

Qy        670 TTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCC 729
Db        241 TGGGCCTCCACGCCGCTCGTCCTCGACGACGCCGACCTCCTCCCCGACGACGACGGCCCC 300

Qy        730 AAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAG 789
Db        301 CAAATCTTCTGGCGCGACGTCGCGGACGCCGTCTCCCGCGTCCTCGCCGCGCACCCGGGC 360

Qy        790 GAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCC 849
Db        361 CCGTTCCGGTGCGTCCGCCTCACGAACGCCTGCAGCTACGCGGGCTCCCGCGGCGCCGCC 420

Qy        850 GTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAG 909
Db        421 GCGCTCTCGCGCTGGCTCGGCGTCCTCGCCGCCAAGGGCGTCCAGGAGCTCGTCCTCGTC 480

Qy        910 CTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGC 969
Db        481 TTCCTCCACGCCTGGCCGGTACGGGTCGAGCCCCCCGCCGATGTCCTCCGCGTCTCCTCC 540

Qy        970 CTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCC 1029
Db        541 CTCCGCCGCCTCTACCTTGGCCTCTGGAGGTCGTTCCCCGACACCGAGCACCTCCGCCCC 600

Qy       1030 CTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACC 1089
Db        601 GGCGCCGACGTCTTCCCCAACCTCGTCGAGCTCGGCATCTGCCGCACCGACATCAAGGCC 660

Qy       1090 CCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATC 1149
Db        661 AAGGACCTCGACCGCCTGCTCCAGTGCAGCCCCGCGCTGGAGACGCTCGCGTTCGTCGTC 720
```

```
Qy       1150 TTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTG 1209

Db        721 AGCTACAACACGCAGCCCAACGTCCGCGTCCGCAGCCGCAGCCTCCGCTGCGTGCTCTTC 780

Qy       1210 ATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGC 1269

Db        781 TGGATGTCCATCGCGGAGGAGCTCGCCGTGGTCGTCGCCCCGCGGCTCGAGCGGCTCATC 840

Qy       1270 GCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAG 1329

Db        841 CTGTGGAACGGGTGCCCCTGCACCCGCCGTGTCCGCCGTGATTTCCGCACCAGGGTCAAG 900

Qy       1330 A 1330

Db        901 A 901
```

```
RESULT 5
AY103647
LOCUS       AY103647                2598 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO142084 mRNA sequence.
ACCESSION   AY103647
VERSION     AY103647.1  GI:21206725
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2598)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2598)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..2598
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:638378"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

  Query Match            3.7%;  Score 90.8;  DB 3;  Length 2598;
  Best Local Similarity   43.3%;  Pred. No. 4e-06;
  Matches  719;  Conservative    0;  Mismatches  909;  Indels   33;  Gaps    5;
```

```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db        351 GCAGCGAGTGGATCCTCGACTGCATCGCCCACGGCGGCGACATCTACGGCGTCACCACCG 410

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db        411 GCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGGCCCGCGCTCCAGGTCGAGCTGC 470

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCG 349

Db        471 TCAGGCATCTCAACGCCGGAATCTTCGGCACCGGCAGCGACGGGCACACGCTGCCGTCGG 530

Qy        350 CCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGA 409

Db        531 AGGTCACCCGCGCGGCGATGCTGGTGCGCATCAACACCCTCCTCCAGGGCTACTCCGGCA 590

Qy        410 GCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCG 469

Db        591 TCCGCTTCGAGATCCTCGAGGCCATCACGAAGCTGCTCAACACCGGTGTCAGCCCCTGCC 650

Qy        470 TGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCG 529

Db        651 TGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTGGTCCCGCTCTCCTACATCGCCG 710

Qy        530 GCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCA 589

Db        711 GCCTCATCACGGGCCGCCCCAACGCGCAGGCCGTCACCG---TCGACGGGAGGAAGGTGG 767

Qy        590 TCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCC 649

Db        768 ACGCCGCCGAGGCGTTCAAGATCGCCGGCATCGAGGGCGGCTTCTTCAAGCTCAACCCCA 827

Qy        650 AGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGA 709

Db        828 AGGAGGGCCTCGCCATCGTCAACGGCACGTCCGTGGGCTCCGCGCTCGCGGCCACCGTGA 887

Qy        710 AGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGG 769

Db        888 TGTACGACGCCAACGTCCTGGCCGTCCTGTCGGAGGTCCTGTCCGCCGTCTTCTGCGAGG 947

Qy        770 CCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCG 829

Db        948 TCATGAACGGCAAGCCCGAGTACACGGACCACCTGACCCACAAGCTGAAGCACCACCCGG 1007

Qy        830 AGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCA 889

Db       1008 GGTCCATCGAGGCCGCGGCCATCATGGAGCACATCCTGGATGGCAGCTCCTTCATGAAGC 1067

Qy        890 AGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGG 949

Db       1068 AGGCCAAGAAGTTGAACGAGCTGGACCCGC-----------TGCTGAAGCCCAAGCAGG 1115

Qy        950 GCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCG 1009

Db       1116 ACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTGGGCCCCCAGATCGAGGTCATCCGCG 1175

Qy       1010 ACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCC 1069

Db       1176 CCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTCATCGACG 1235

Qy       1070 CCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCT 1129

Db       1236 TCCACCGCGGCAAGGCGCTGCACGGCGGCAACTTCCAGGGCACCCCCATCGGCGTGTCCA 1295

Qy       1130 GGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCG 1189

Db       1296 TGGACAACGCCCGCCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCG 1355

Qy       1190 CCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGA-- 1247

Db       1356 AGCTCGTCAACGAGTTCTACAACAACGGGCTCACCTCCAACCTGGCCGGCAGCCGCAACC 1415

Qy       1248 ----CCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1303

Db       1416 CCAGCCTGGACTACGGCTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCCGAGC 1475

Qy       1304 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1363
```

```
Db       1476 TCCAGTACCTGGGCAACCCCATCACCAACCACGTGCAGAGCGCGGAGCAGCACAACCAGG 1535

Qy       1364 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1423

Db       1536 ACGTGAACTCCCTGGGCCTCGTCTCGGCCAGGAAGACCGCCGAGGCGATCGACATCCTGA 1595

Qy       1424 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1483

Db       1596 AGCTCATGTCGTCCACCTACATCGTGGCGCTGTGCCAGGCCGTGGACCTGCGCCACCTCG 1655

Qy       1484 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1543

Db       1656 AGGAGAACATCAAGGCGTCGGTGAAGAACACCGTGACCCAGGTGGCCAAGAAGGTGCTGA 1715

Qy       1544 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAA---------CCGCG 1594

Db       1716 CCATGAACCCCTCGGGCGAGCTCTCCAGCGCCCGCTTCAGCGAGAAGGAGCTGATCAGCG 1775

Qy       1595 AGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCG 1654

Db       1776 CCATCGACCGCGAGGCCGTGTTCACGTACGCGGAGGACGCGGCCAGCGGCAGCCTGCCGC 1835

Qy       1655 AGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGA 1714

Db       1836 TGATGCAGAAGCTGCGCGCCGTGCTGGTGGACCACGCCCTCAGCAGCG---GCGACGCGG 1892

Qy       1715 ACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGC 1774

Db       1893 AGCGGGAGCCCTCCGTGTTCTCCAAGATCACCAGGTTCGAGGAGGAGCTCCGCGCGGTGC 1952

Qy       1775 TGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTG 1815

Db       1953 TGCCCCAGGAGGTGGAGGCCGCCCGCGTGGCGGTCGCCGAG 1993
```

```
RESULT 6
CL961989
LOCUS       CL961989                1398 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC006991 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL961989
VERSION     CL961989.1  GI:52378720
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1398)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1398
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN
```

```
  Query Match             3.5%;  Score 86.6;  DB 9;  Length 1398;
  Best Local Similarity   45.9%;  Pred. No. 2.2e-05;
  Matches  508;  Conservative    0;  Mismatches  569;  Indels   30;  Gaps    5;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Db        166 GTCGAGAGGCACGTCGGCGAGGCCGACATCGTCTTCGTCTCCGTGAACACCCCCACCAAG 225

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Db        226 GCCCGCGGCCTCGGCGCCGGCAAGGCCGCCG---ACCTCACCTACTGGGAGAGCGCGGCG 282

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Db        283 CGGATGATCGCCGCCGTGGCGACGTCCGACAAGGTCGTCGTCGAGAAGTCCACCGTGCCG 342

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Db        343 GTCAAGACCGCGGAGGCCATCGAGAAGATCCTCGACCACAACGGCCGCGACGGCGTCGGC 402

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Db        403 TTCCAGATCCTCTCCAACCCGGAGTTCCTCGCCGAGGGCACCGCCATCCGCGACCTGCTC 462

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Db        463 GCCCCCGA---------CCGCGTCCTCATCGGCGGCCGCGAGACCGCCGCGGGGCGCGCC 513

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Db        514 GCCGTCCAGGCGCTCAAGGACGTGTACGCCCGATGGGTCCCCGAGGAGAGGATC---CTC 570

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Db        571 ACCACCAACCTCTGGTCCGCCGAGCTGTCCAAGCTCGCCGCCAACGCGTTCCTGGCCCAG 630

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Db        631 AGGATCTCGTCCGTCAACGCCATGTCCGCGCTCTGCGAGGCCACCGGCGCCGACGTCGCC 690

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGA---GATCTGCGGCAAGAAGGCCATCGGCACC 597

Db        691 GAGGTCGCCTACGCCGTGGGCAAGGACTCCAGGATCGGCGCCAAGTTTCTCAACGCCAGC 750

Qy        598 GTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGC 657

Db        751 GTGGGATTCGGTGGCTCCTGCTTCCAGAAAGACATCCTCAACCTGCTG------------ 798

Qy        658 TGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGC 717

Db        799 TACATCTGCGAGTGCAATGGCCTCCCCGAGGTGGCCAACTACTGGAAGCAGGTTATCAAG 858

Qy        718 ATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACC 777

Db        859 ATCAACGACTACCAGAAGAGCCGGTTCGTGAACCGCGTGGTGTCGTCCATGTTCAACACC 918

Qy        778 GCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCC 837

Db        919 GTGGCCGGCAAGAAGATCGCCGTGCTGGGCTTCGCCTTCAAGAAGGACACCGGCGACACC 978

Qy        838 TACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTG 897

Db        979 AGGGAGACGCCCGCCATCGACGTCTGCAAGGGGCTGATCGGCGACAAGGCCAAGGTGAGC 1038

Qy        898 GACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC 957

Db       1039 ATCTACGACCCCCAGGTGACGGAGGACCAGGTCCAGCGCGACCTCGCCATGAGCAAGTTC 1098

Qy        958 CACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTAC 1017

Db       1099 GATTGGGATCACCCGGTCCATCTCCAGCCGATGAGCCCCACGGCCATCAAGCAGGTGTCC 1158

Qy       1018 TTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATC 1077
```

```
Db        1159 GTCGCCTGGGACGCCTACGAGGCGGCCAGGGACGCCCACGGCGTCTGCATCCTCACCGAG 1218

Qy        1078 AACAACGAGACCCCCGGCATCCGCTAC 1104

Db        1219 TGGGACGAGTTCAGGAGCCTCGACTAC 1245


RESULT 7
BM321451
LOCUS       BM321451                 951 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321451
VERSION     BM321451.1  GI:18055857
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 951)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
   MEDLINE  21819461
    PUBMED  11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 951   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..951
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.5%;  Score 85.6;  DB 4;  Length 951;
  Best Local Similarity   46.0%;  Pred. No. 3.2e-05;
  Matches  323;  Conservative    0;  Mismatches  376;  Indels    3;  Gaps    1;

Qy        1734 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1793

Db         219 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 278

Qy        1794 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1853

Db         279 CATCGCGCAGGCCGGCGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 338

Qy        1854 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1913

Db         339 GGAGTGCACCTGGAAGGCCCTGTGCTTCGGCCCCTACCAGGGCCCTCAGATCATCGTCGA 398

Qy        1914 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1973

Db         399 CGACGGCGGTGACGCGACTCTGATGATCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 458

Qy        1974 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2033

Db         459 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 518

Qy        2034 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2093

Db         519 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 578

Qy        2094 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2153

Db         579 CGAGGAGACGACGACTGGCGTNATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 638

Qy        2154 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2213

Db         639 GTTCCCGGCCGTCAACGTCAACGAC---TCTNTCACCAAGAGCAAGTTCGACAACATCTA 695

Qy        2214 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2273

Db         696 CGGCTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGG 755

Qy        2274 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2333

Db         756 CAAGGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCG 815

Qy        2334 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393

Db         816 CGGCCAGGGCTGCCGCGTCATCGTGACGGAGATCGACCCCATCTGCGCGCTGCAGGCGTC 875

Qy        2394 CATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGA 2435

Db         876 GATGGCCGGCTTCGAGGTCAACACGCTCGAGGCGGGGCTCGA 917


RESULT 8
CK159167/c
LOCUS       CK159167                 869 bp    mRNA    linear   EST 05-DEC-2003
DEFINITION  FGAS040564 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA,
            mRNA sequence.
ACCESSION   CK159167
VERSION     CK159167.1  GI:38985053
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
            Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
            Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
            Penniket,C., Roach,J.L. and Sarhan,F.
  TITLE     Functional Genomics of Abiotic Stress In Wheat and Canola Crops
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [128,636].
            Plate: TaLt537  row: N  column: 23.
FEATURES             Location/Qualifiers
     source          1..869
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Wheat line PI 178383"
                     /db_xref="taxon:4565"
                     /lab_host="DH5 alpha"
                     /clone_lib="Triticum aestivum FGAS: TaLt5"
                     /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
                     subtractive hybridization) cDNA library from genotype
                     PI178383 cold hardened at 2 C for 21 days and 49 days
                     (equal amount of cDNA pooled together before subtraction,
                     tester) and subtracted against genotype Norstar cold
```

```
                hardened at 2 C for 1 day (24 H)(driver). Modified Smart
                cDNA (Clontech)priming and non-directional cloning"
ORIGIN

  Query Match              3.4%;  Score 85;  DB 7;  Length 869;
  Best Local Similarity   45.4%;  Pred. No. 4.1e-05;
  Matches  304;  Conservative    0;  Mismatches  365;  Indels    0;  Gaps    0;

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
Db        827 ATGGCGGCGGCCCGCACAACGAGGATCACCAACCAGGACGAGAACAACCACAACCACCAC 768

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
Db        767 CACAACGACGACAACAGCGACACAACGACCACCAACAACACCAACAACAACAACAACAAC 708

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
Db        707 ACCAACAACAACAACAACAACAACAACAACAACGACAACCACCACCACAACCACAACAAC 648

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
Db        647 AACAACAACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 588

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
Db        587 AACAACAACAACAACAACAACAACAACAACAACAACAACGACAACAACAACAACAACAAC 528

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
Db        527 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 468

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
Db        467 AACAACAACAACAACAACAACAACAACGACGACGACGACAACGACGACGACGACGACGAC 408

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
Db        407 GACAACAACGACAACAACGACAACAACAACAACGACAACAACAACAACAACAACAACAAC 348

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
Db        347 GACCACAACGACAACGACAACAACAACAACAACAACAACAACAACGACAACAACCACAAC 288

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
Db        287 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 228

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTG 1239
Db        227 AACAACAACAACAACAACGACGACAACGACGTCAACGATAACAACGACAACAACGACAAC 168

Qy       1240 GGCAGCGAC 1248
Db        167 AACGGCAAC 159


RESULT 9
CL973991
LOCUS       CL973991            3069 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC025416 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL973991
VERSION     CL973991.1  GI:52402507
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 3069)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..3069
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match              3.4%;  Score 84;  DB 9;  Length 3069;
  Best Local Similarity   42.8%;  Pred. No. 6.7e-05;
  Matches  536;  Conservative    0;  Mismatches  710;  Indels    6;  Gaps    2;

Qy         37 ACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAG 96
Db        625 AGCGGCATCGGCAAGACGACGCTGGCGAGCTACCTCTACCGGAGGAACAGCATCCGGAAG 684

Qy         97 TGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAG 156
Db        685 CACTTCGCCTGCACGGCGTGGGTCCACGTCCCCCGCAAGTTCCGCTTCGCCGACCTCCTC 744

Qy        157 GGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCC 216
Db        745 CACGACGTCATCCGCCAGGCCAACGCCTCCTCCGGCGACGACGAGCTCGGCGGCGAGGC- 803

Qy        217 AACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAG 276
Db        804 --CGCCGACGACGACGTGCGGCGCCGGGGTGGGATGGAGGCGGAGAGGCTCGTCAAGGCG 861

Qy        277 CAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGC 336
Db        862 CGCCTCGCCGCCGCGCTCCAGGGGAAGAGGTACCTCGTGGTGCTCGACGACGTCCGGAGC 921

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAG 396
Db        922 AGGGAGGAGTGGGCCTTCTTCCTCGCCGCGCTCCCCGGCGGCCTCGCCGGCAGCTGCGTG 981

Qy        397 CGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGC 456
Db        982 CTCGTCACGACGCAGGTCAAGATCAGCGAGGAGTCGGAGCAACCCGTGGCCGCCGGCGGC 1041

Qy        457 GCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATC 516
Db       1042 GGCGTCGTGCGCACCAGGGAGCTCGGGAAGCTGCGGCCGGAGCAGGCCGCCGAGCTGTTC 1101

Qy        517 GGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGC 576
Db       1102 CGCCGGAGGGTGTACGGCCACGGCGAGCCGGAGCCGAAGAAGATGGAGCAGCTCAAGAGC 1161

Qy        577 GGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGC 636
Db       1162 CTCGTGGACTCCATGACGAAGGGCTCCAACCTGCCGCTCAACATCGTCATGCTCGCCGGA 1221

Qy        637 AACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTG 696
Db       1222 CTCTTGAGGTCCAAGAAGGAGGATGAATGGGAGGCCGTGATCCATAGCCTCGATGATACG 1281

Qy        697 CCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAG 756
Db       1282 CCACCCCCGACATCGCCGGAGCAGACGCCGCCGGAGGCGGAGCCGAAGAAGGACAGCCTC 1341
```

```
Qy     757 GAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACC 816
Db    1342 GATGATAGCGAGACGGCGCCGACATCGACGGAGCAAACGAAGAAGACGAAGAAGAGGAAG 1401

Qy     817 AAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGC 876
Db    1402 CAGATGGCGACATCGATGGACAAGATACTGACGGTGTGCATGGACGACCTGCCGACGCAT 1461

Qy     877 ACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936
Db    1462 CTGAAGCCGTGCTTCCTCTACTTCGCCGGGTTCACCGCGCAGACGCCGATCTGCGCCGGG 1521

Qy     937 GAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTG 996
Db    1522 AAGCTGGTCCGGCTGTGGGTCGCCGAGGGGTTCATGCAGGCGAAGAACGGGCAGACGGTG 1581

Qy     997 CTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACC 1056
Db    1582 GAGGAGCACGGCGAGGAGTGCCTCAAGGAGCTCATCTCCCGGTGCCTCGTCCAGCTCGTC 1641

Qy    1057 GCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTG 1116
Db    1642 GAGACGGACGCCGGCGGCGGCAGGGTCGCCGCCGTCAGCATCCACCAAGCGGTGCTCGAC 1701

Qy    1117 CTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCA---GAGCAGCATGACCAAGATC 1173
Db    1702 TTCGTCCAGGCGGAGGCGCGCGACACCAACTTCCTCCACGTGCACAGCGGCGCGGCGGGC 1761

Qy    1174 CTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTG 1233
Db    1762 CTCTCCAACGGCGCCGCCCGCCGCCTCGCGCTCCGCAACACCTACGACACCGACCTCGCC 1821

Qy    1234 TACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGC 1285
Db    1822 GTCATGCTCGAGGCGCCCAAGCTCCACACCCTCCTCTGCGACATCCCGGAGC 1873
```

```
RESULT 10
CL971508
LOCUS       CL971508                1941 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC021485 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL971508
VERSION     CL971508.1  GI:52397596
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1941)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1941
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.4%;  Score 83.8;  DB 9;  Length 1941;
  Best Local Similarity   42.9%;  Pred. No. 7e-05;
  Matches  769;  Conservative    0;  Mismatches  992;  Indels   30;  Gaps    6;
```

```
Qy     267 CCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGA 326
Db     165 CGCCGCCAAGAACCAGGTCGCCATGAACCCAACCAACACCGTCTTCGATGCGAAGCGGTT 224

Qy     327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCAC 386
Db     225 GATTGGTAGGAGGTTCTCCGACCCGTCCGTGCAGAGTGACATGAAGCTATGGCCGTTCAA 284

Qy     387 CCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCT 446
Db     285 GGTGGTGCCTGGCCCCGGCGACAAGCCGATGATCGTCGTCCAGTACAAGGGCGAGGAGAA 344

Qy     447 GGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCC 506
Db     345 GCAGTTCGCCGCGGAGGAGATCTCGTCCATGGTGCTGATCAAGATGAGGGAGATCGCGGA 404

Qy     507 CAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGAT 566
Db     405 GGCCTACCTCGGCTCGTCCATCAAGAACGCCGTGGTCACCGTCCCGGCCTACTTCAACGA 464

Qy     567 CGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACAT 626
Db     465 CTCGCAGAGGCAGGCCACCAAGGACGCCGGCGTCATCGCCGGGCTCAACGTGATGCGCAT 524

Qy     627 CATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCAT 686
Db     525 CATCAACGAGCCCACCGCCGCCGCCATCGCGTACGGCCTCGACAAGAAGGCGACGAGCAG 584

Qy     687 CGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCC 746
Db     585 CGGCGAGAAGAACGTGCTCATCTTTGACCTCGGCGGCGGCACGTTCGATGTGTCGCTGCT 644

Qy     747 CCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGG 806
Db     645 CACCATCGAGGAGGGCATCTTCGAGGTCAAGGCCACCGCCGGCGACACCCACCTCGGCGG 704

Qy     807 CAAGATCACCAAGATCGGCCCCGAGAACCCCTAC---AACACCCCCGTGTTCGCCATCAA 863
Db     705 CGAGGACTTCGACAACCGCATGGTGAACCACTTCGTGCAGGAGTTCAAGCGCAAGAACAA 764

Qy     864 GAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCAC 923
Db     765 GAAGGACATCAGCGGCAACCCGCGCGCGCTGCGGCGGCTCAGGACGGCGTGCGAGCGCGC 824

Qy     924 CCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAA 983
Db     825 CAAGCGCACGCTGTCGTCGACGGCGCAGACCACCATTGAGATCGACTCGCTGTACGAGGG 884

Qy     984 GAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTT 1043
Db     885 CATCGACTTCTACACGACGATCACCAGGGCGCGGTTCGAGGAGCTCAACATGGACCTGTT 944

Qy    1044 CCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTA 1103
Db     945 CCGCAAGTGCATGGAGCCCGTGGAGAAGTGCCTCCGCGACGCCAAGATGGACAAGAGCAG 1004

Qy    1104 CCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCAT 1163
Db    1005 CGTGCACGACGTCGTGCTC---------GTCGGCGGCTCCACCCGCATCCCCAAGGTGCA 1055

Qy    1164 GACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACAT 1223
Db    1056 GCAGCTGCTCCAGGACTTCTTCAACGGCAAGGAGCTCTGCAAGAGCATCAACCCCGACGA 1115

Qy    1224 GGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGA 1283
```

```
Db     1116 GGCCGTGGCGTACGGCGCCGCCGTCCAGGCCGCCATCCTGAGCGGCGAGGGCAACGAGAA 1175

Qy     1284 GCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGA 1343
Db     1176 GGTGC---AGGACCTGCTCCTGCTCGACGTCACCCCGCTGTCCCTCGGCCTCGAGACCGC 1232

Qy     1344 GCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC---GTGCAGCC 1400
Db     1233 CGGCGGCGTCATGACCGTGCTCATCCCGAGGAACACCACCATCCCGACCAAGAAGGAGCA 1292

Qy     1401 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1460
Db     1293 GGTCTTCTCCACCTACTCCGACAACCAGCCCGGCGTCCTGATCCAGGTGTACGAGGGCGA 1352

Qy     1461 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA------- 1513
Db     1353 GCGCGCCAGGACCAAGGACAACAACCTCCTCGGCAAGTTCGAGCTCTCCGGCATCCCACC 1412

Qy     1514 --AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGA 1571
Db     1413 GGCGCCGCGCGGCGTGCCCCAGATCAACGTCTGCTTCGACATCGACGCCAACGGCATCCT 1472

Qy     1572 GCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGA 1631
Db     1473 GAACGTGTCCGCGGAGGACAAGACGACGGGGCAGAAGAACAAGATCACCATCACCAACGA 1532

Qy     1632 CCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCA 1691
Db     1533 CAAGGGGCGGCTGAGCAAGGAGGAGATCGAGAAGATGGTGCAGGAGGCGGAGAAGTACAA 1592

Qy     1692 GATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGC 1751
Db     1593 GGCCGAGGACGAGGAGCACAAGAAGAAGGTGGACGCCAAGAACGCCCTCGAGAACTACGC 1652

Qy     1752 CCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCAT 1811
Db     1653 CTACAACATGCGCAACACCATCAAGGACGAGAAGATCGCCTCCAAGCTCGCCGCCGACGA 1712

Qy     1812 CGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGAC 1871
Db     1713 CAAGAAGAGGATCGAGGACGCCATCGATGGCGCCATCAGCTGGCTCGACACCAACCAGCT 1772

Qy     1872 CTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCC 1931
Db     1773 CGCCGAGGCCGACGAGTTCGAGGACA---AGATGAAGGAGCTGGAGGGCATCTGCAACCC 1829

Qy     1932 CCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTT 1991
Db     1830 GATCATCGCCAAGATGTACCAGGGCCCCGGCGCCGACATGGCCGGAGGCATGGACGAGGA 1889

Qy     1992 CTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTA 2042
Db     1890 CGCGCCGGCGGGCGGCAGCGGCGCAGGCCCCAAGATCGAGGAGGTCGACTA 1940
```

RESULT 11
BM320864

```
LOCUS       BM320864          1132 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM320864
VERSION     BM320864.1  GI:18055270
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1132)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1132   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..1132
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.2%;  Score 80.2;  DB 4;  Length 1132;
  Best Local Similarity   45.3%;  Pred. No. 0.0003;
  Matches  375;  Conservative    0;  Mismatches  444;  Indels    9;  Gaps    2;

Qy      139 CGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGAC 198
Db       79 CGCCGCGAGGGCAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAG 138

Qy      199 TGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCC 258
Db      139 TACAACAGCCCCAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTC 192

Qy      259 CGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG 318
Db      193 TGCCAGATCGCCTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCAC 252

Qy      319 CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCC 378
Db      253 GAGCTCACCCGCTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACT 312

Qy      379 CAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAG 438
Db      313 GGCCTGCTGCTGGCCCGCCGTGTGCTGAAGAAGCTCAACCTCGACTCCAAGTACGAGGGT 372

Qy      439 GCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAG 498
Db      373 GTCAAGAAGGTCAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGT 432

Qy      499 TGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAG 558
Db      433 TTCAAGGCCCTGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCC 492

Qy      559 ATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCC 618
Db      493 GCCCTCAAGGGCATGTGCGACGGCGGCGTCAACGTCCCC---CACAGCGAGACCCGCTTC 549

Qy      619 GTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATC 678
Db      550 GTCGGCTTCAACGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTC 609

Qy      679 AGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAG 738
Db      610 GGCGGCCACGTCGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGAC 669

Qy      739 CAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAG 798
Db      670 CGCCAGTTCTCGCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTAC 729

Qy      799 AAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCC 858
Db      730 ACCGAGGCCCACAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAG 789

Qy      859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918
```

```
Db        790 CCCGAGGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAG 849

Qy        919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966

Db        850 AACCGCGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCC 897
```

RESULT 12
BM320900

```
LOCUS       BM320900                1165 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM320900
VERSION     BM320900.1  GI:18055306
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1165)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1165   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..1165
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match              3.2%;  Score 79.6;  DB 4;  Length 1165;
  Best Local Similarity   45.1%;  Pred. No. 0.00039;
  Matches  428;  Conservative    0;  Mismatches  506;  Indels   16;  Gaps    3;

Qy         90 CATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCG 149

Db         46 CGTCAAGAACAAGGCGTACTTCAAGCGCTTCCAGACCCAGTTCCGTCGCCGCCGCGAGGG 105

Qy        150 CAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCG 209

Db        106 CAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAGTACAACAGCCC 165

Qy        210 CCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCC 269

Db        166 CAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTCTGCCAGATCGC 219

Qy        270 CAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAA 329

Db        220 CTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCACGAGCTCACCCG 279

Qy        330 CCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCT 389

Db        280 CTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACTGGCCTGCTGCT 339

Qy        390 GTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGA 449

Db        340 GGCCCGCCGTGTGCTGAAGAAGCTCAACCTCTACTCCAAGTACGAGGGTGTCAAGAAGGT 399

Qy        450 CACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAA 509

Db        400 CAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGTTCAAGGCCC-- 457

Qy        510 GATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGA 569

Db        458 --TGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCCGCCCTCAA 515

Qy        570 GATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCAT 629

Db        516 GGGCATGTGCGACGGCGGCGTCAACGTCCCCCACAGCGAGACCCGCTTCGTCGGCTTCAA 575

Qy        630 CGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGA 689

Db        576 CGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTCGGCGGCCACGT 635

Qy        690 GACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCT 749

Db        636 CGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGACCGCCAGTTCTC 695

Qy        750 GACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAA 809

Db        696 GCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTACACCGAGGCCCA 755

Qy        810 GATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAA 869

Db        756 ------CAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAGCCCGA 809

Qy        870 GGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGA 929

Db        810 GGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAGAACCG 869

Qy        930 CTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGT 989

Db        870 CGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCCAAGAGCCAGTAAATTCC 929

Qy        990 GACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGG 1039

Db        930 CGGTGCCCTGTACACCGCCATCTCCCGCTGCGGCGTTGCGCTGCTNGCGG 979
```

RESULT 13
BM321430

```
LOCUS       BM321430                 867 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
            sequence.
ACCESSION   BM321430
VERSION     BM321430.1  GI:18055836
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
  MEDLINE   21819461
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 867   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..867
                     /organism="Mastigamoeba balamuthi"
```

```
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match              3.2%;  Score 79;  DB 4;  Length 867;
  Best Local Similarity   45.7%;  Pred. No. 0.00049;
  Matches  314;  Conservative    0;  Mismatches  370;  Indels    3;  Gaps    1;

Qy         58 CAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAG 117

Db        131 CCGCACAAGATGCGCGAGTGCCTGCCGGTCATCATCCTGCTGCGCAACAGGTTGAAGTAC 190

Qy        118 GGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAG 177

Db        191 GCGCTGACCCGCCGTGAGGTCACCTCGATCGTGATGCAGCGCCTGATCAAGATCGACGGC 250

Qy        178 GAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTG 237

Db        251 AAGGTCCGCACCGACACGCACTTCCCCGCGGGCTTCATGGACGTCGTCTCGATCGACAAG 310

Qy        238 GCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCC 297

Db        311 ACCGACGAGCACTTCCGCCTGCTCTACGACACCAAGGGCCGCTTCCAGGCGCACCGCATC 370

Qy        298 ACCAGCCGCGAGCTGCAGGT---GCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAG 354

Db        371 AACTCGACCGAGGCCAAGTTCAAGCTCGGCAAGGTCCGCCGCGTGCAGCTCGGCAACAAG 430

Qy        355 CGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATC 414

Db        431 GGCATCCCGTACCTGGTGACCCACGACGCGCGCACGATCCGCTACCCCAACCCCGACATC 490

Qy        415 AAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTG 474

Db        491 AAGGTCAACGACACGGTCAAGATCGACCTGCGCTCGGGCAAGATCATCGACTTCGTCAGG 550

Qy        475 GAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTC 534

Db        551 TTCGAGATCGGCAACCTCGTCATGATCACTGGCGGACGCAACCTTGGCCGCGTCGGCGTC 610

Qy        535 ATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGC 594

Db        611 ATTGTGCGCCGCGAGAAGCACGAGGGCTCGTTCGAGATCATCCACGTCAAGGACGCCGTC 670

Qy        595 ACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654

Db        671 GGCCACCAGTTCGCGACGCGCCTGACCAACGTCTTCGTGATCGGCAAGGGCACCAAGTCC 730

Qy        655 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCC 714

Db        731 CTCGTCACGCTGCCCGCCGGCAAGGGCATCAAGAAGTCGATCATCGAGGAGTTCCAGGCG 790

Qy        715 GGCATGGACGGCCCCAAGGTGAAGCAG 741

Db        791 CGCCACGCGCACAGGACCGAGCAGGAG 817


RESULT 14
CL962901
LOCUS       CL962901               11691 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC008219 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL962901
VERSION     CL962901.1  GI:52380530
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 11691)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..11691
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match              3.2%;  Score 78.4;  DB 9;  Length 11691;
  Best Local Similarity   42.2%;  Pred. No. 0.00074;
  Matches  567;  Conservative    0;  Mismatches  771;  Indels    6;  Gaps    2;

Qy        150 CAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCG 209

Db       2991 CAAGAAGGCCTTCGACAAGCAGGGCCTGAAGGTCGAGCTGGGCGTGAAGATCGGCGAAGT 3050

Qy        210 CCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCC 269

Db       3051 GAAGACGGGCGGCGACGGCGTGACCGTGGCCTACACCGATGCCAAGGGCGAGGCCCGCGA 3110

Qy        270 CAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAA 329

Db       3111 ACTGGCGGTGGACAAGCTCATCGTCTCCATCGGCCGCGTGCCCAACACCATCGGCCTCAA 3170

Qy        330 CCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCT 389

Db       3171 CCCCGAAGCCGTGGGCCTGAAGCTCGACGAGCGCGGCGCCATCGTGGTCGATGCCGACTG 3230

Qy        390 GTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGG- 448

Db       3231 CAAGACCAACCTGCCCGGCGTGTGGGCGGTGGGCGACGTGGTGCGCGGCCCGATGCTGGC 3290

Qy        449 --ACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCC 506

Db       3291 GCACAAGGCCGAGGAAGAGGGCGTGGCGGTGGCCGAGCGCATCGCGGGCCAGCATGGCCA 3350

Qy        507 CAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGAT 566

Db       3351 CGTGAACTTCAACACCATCCCCTGGGTCATCTACACCAGCCCCGAGATCGCCTGGGTGGG 3410

Qy        567 CGAGATCTGCGGCAAG---AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAA 623

Db       3411 CCGCACCGAGCAGCAGCTCAAGGCCGACGGCGTGCAGTACAAGGCCGGCACCTTCCCGTT 3470

Qy        624 CATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCC 683

Db       3471 CCTGGCGAACGGCCGCGCCCGTGCGCTGGGCGACACGACCGGCATGGTCAAGTTCCTGGC 3530

Qy        684 CATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTG 743

Db       3531 CGATGCGGCCACGGACGAGATCCTGGGCGTGCACATCGTGGGCCCGATGGCCAGCGAGCT 3590

Qy        744 GCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGA 803

Db       3591 GATCTCCGAGGCCGTGGTGGCCATGGAGTTCAAGGCCAGCGCCGAGGACATCGCGCGCAT 3650
```

```
Qy        804 GGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAA 863
Db       3651 CTGCCACGCCCATCCCTCGCTTTCCGAGGCGACCAAGGAAGCGGCCCTGGCCGCGTATGA 3710

Qy        864 GAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCAC 923
Db       3711 GGCCGAGCTGGCCGCCAAGGGGTTCCGCAGCGACCCGGCCCAGCTGCGCGCCGTCGAGGC 3770

Qy        924 CCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAA 983
Db       3771 GCTGCAGCGCTGCGCGGACGACTGGGCCGCCTACAAGGAAAAGCGCTCCAACGCGCTCAA 3830

Qy        984 GAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTT 1043
Db       3831 GAAGCTCATCCACCGCCCCGAGGTGCCGCGCGGCGTGTACATGTACGGGGGCGTGGGCCG 3890

Qy       1044 CCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTA 1103
Db       3891 TGGCAAGAGCTTCCTGATGGACTGCTTCTTCAACGCCGTGCCGATCCGCCGCAAGGTGCG 3950

Qy       1104 CCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCAT 1163
Db       3951 GCTGCATTTCCACGAATTCATGCGTGAGGTGCACCGCGAACTGGCGGGCCTGCAGGGCAC 4010

Qy       1164 GACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACAT 1223
Db       4011 CGTGAACCCGCTGGACGTGCTGGGCGAGCGCATCGCCAAGCGCTACAAGCTCATCTGCTT 4070

Qy       1224 GGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGA 1283
Db       4071 CGACGAATTCCACGTGGCCGACATCACGGACGCCATGATCCTGCACCGCCTGCTGGCGGC 4130

Qy       1284 GCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGA 1343
Db       4131 GCTGTTCGACAACGGCGTGGGCTTCGTCACCACCTCGAACTTCAAGCCCGACGACCTGTA 4190

Qy       1344 GCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCAT 1403
Db       4191 CCCGGGCGGCCTGCACCGCGACCGCATCCTGCCGGCCATCGATCTGCTCAACGAGAAGCT 4250

Qy       1404 CGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCT 1463
Db       4251 GGAGGTGGTGAACGTGGACAACGGCACCGACTACCGGCGCCGCACGCTGGAGCAGGTCCG 4310

Qy       1464 GAACTGGGCCAGCCAGATCTACCC 1487
Db       4311 GCTCTACCACTGCCCGCTGGGCCC 4334
```

```
RESULT 15
CL961326
LOCUS       CL961326               1962 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC005994 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL961326
VERSION     CL961326.1  GI:52377391
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1962)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1962
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.2%;  Score 78.2;  DB 9;  Length 1962;
  Best Local Similarity   41.6%;  Pred. No. 0.00071;
  Matches  737;  Conservative    0;  Mismatches 1013;  Indels   21;  Gaps    3;
```

```
Qy        291 CAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGC 350
Db        162 CATCGGCGACGCCGCCAAGAACCAGGTCGCCATGAACCCCACCAACACCGTCTTTGATGC 221

Qy        351 CGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAG 410
Db        222 AAAGCGGCTGATCGGGCGGCGCTTCTCGGACCCGTCGGTGCAGGCGGACATGAAGATGTG 281

Qy        411 CATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGT 470
Db        282 GCCGTTCAAGGTGGTGCCTGGCCCAGCCGACAAGCCGATGATCGTGGTGACGTACAAGGG 341

Qy        471 GCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGG 530
Db        342 CGAGGAGAAGAAGTTCTCGGCGGAGGAGATCTCGTCCATGGTGCTCACCAAGATGAAGGA 401

Qy        531 CTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCAT 590
Db        402 GATCGCCGAGGCCTTCCTCAGCACCACCATCAAGAACGCCGTCATCACCGTCCCGGCCTA 461

Qy        591 CGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCA 650
Db        462 CTTCAACGACTCCCAGCGCCAGGCCACCAAGGACGCCGGCGTCATCTCCGGCCTCAACGT 521

Qy        651 GCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAA 710
Db        522 CATGCGCATCATCAACGAGCCCACCGCCGCCGCCATCGCCTACGGCCTCGACAAGAAGGC 581

Qy        711 GCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGC 770
Db        582 CGCCAGCACCGGCGAGAAGAATGTCCTCATCTTCGACCTCGGCGGCGGCACCTTCGACGT 641

Qy        771 CCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCC-- 828
Db        642 CTCCATCCTCACCATCGAGGAGGGCATCTTCGAGGTCAAGGCCACCGCCGGCGACACCCA 701

Qy        829 -------GAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAA 881
Db        702 CCTGGGAGGCGAGGACTTCGACAACCGGATGGTGAACCACTTCGTGCAGGAGTTCAAGAG 761

Qy        882 GTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTT------CTG 935
Db        762 GAAGCACAAGAAGGACATCACCGGCAACCCGAGGGCGCTCCGGCGGCTGAGGACGGCGTG 821

Qy        936 GGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGT 995
Db        822 CGAGAGGGCGAAGAGGACGCTGTCCTCCACCGCCCAGACCACCATCGAGATCGACTCGCT 881

Qy        996 GCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACAC 1055
Db        882 CTACGAGGGCATCGACTTCTACGCCACCATCACCCGCGCCCGCTTCGAGGAGCTCAACAT 941

Qy       1056 CGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGT 1115
Db        942 GGACCTCTTCCGCCGGTGCATGGAGCCCGTGGAGAAGTGCCTCCGCGACGCCAAGATGGA 1001
```

```
Qy    1116 GCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175

Db    1002 CAAGGCCCAGATCCACGATGTCGTCCTCGTCGGAGGCTCCACCCGTATCCCCAAGGTGCA 1061

Qy    1176 GGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTA 1235

Db    1062 GCAGCTCCTGCAGGACTTCTTCAACGGGAAGGAGCTCTGCAAGAGCATCAACCCCGACGA 1121

Qy    1236 CGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCA 1295

Db    1122 GGCCGTGGCGTACGGCGCCGCCGTCCAGGCCGCCATCCTCAGCGGCGAGGGCAACCAGAG 1181

Qy    1296 CCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCT 1355

Db    1182 GGTGCAGGACCTGCTCCTCCTCGACGTCACGCCGCTCTCGCTCGGGTTGGAGACGGCCGG 1241

Qy    1356 GTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGA 1415

Db    1242 GGGTGTCATGACCGTGCTGATCCCGAGGAACACCACCATCCCAACCAAGAAGGAGCAGGT 1301

Qy    1416 GAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAG 1475

Db    1302 CTTCTCCACCTACTCCGACAACCAGCCCGGCGTGCTCATCCAGGTGTACGAGGGCGAGAG 1361

Qy    1476 CCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGC 1535

Db    1362 GACGAGGACCAAGGACAACAACCTGCTCGGCAAGTTCGAGCTCACCGGCATCCCGCCGGC 1421

Qy    1536 CCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGA 1595

Db    1422 GCCGAGGGGCGTGCCCCAGATCAACGTGACGTTCGACATCGACGCGAACGGCATCCTGAA 1481

Qy    1596 GATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGA------CCCCAGCAAGGACCTGGT 1649

Db    1482 CGTGTCGGCGGAGGACAAGACGACGGGGAAGAAGAACAAGATCACCATCACCAACGACAA 1541

Qy    1650 GGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTT 1709

Db    1542 GGGGCGGCTGAGCAAGGAGGAGATCGAGCGGATGGTGCAGGAGGCCGAGAAGTACAAGGC 1601

Qy    1710 CAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAA 1769

Db    1602 CGAGGACGAGCAGGTGCGGCACAAGGTGGAGGCCCGCAACGCGCTGGAGAACTACGCGTA 1661

Qy    1770 GCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGAC 1829

Db    1662 CAACATGCGCAACACGGTGCGCGACGAGAAGATCGCCTCCAAGCTCCCCGCCGACGACAA 1721

Qy    1830 CCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTG 1889

Db    1722 GAAGAAGATCGAGGACGCCATCGAGGATGCCATCAAGTGGCTCGACGGCAACCAGCTCGC 1781

Qy    1890 GCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTG 1949

Db    1782 CGAGGCCGACGAGTTCGAGGACAAGATGAAGGAGCTCGAGAGCCTCTGCAACCCGATCAT 1841

Qy    1950 GTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGC 2009

Db    1842 CTCAAAGATGTACCAGGGCGGCGCCGGTGGTCCGGCGGGCATGGACGAAGACGCCCCCAA 1901

Qy    2010 CAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Db    1902 CGGCGGTGCCGGCACCGGCGGTGGCAGCGGC 1932
```

Search completed: June 2, 2005, 02:53:19
Job time : 7886.14 secs

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:          June  2, 2005, 03:13:23 ; Search time 34 Seconds
                                           (without alignments)
                                           10815.332 Million cell updates/sec

Title:           US-09-610-313B-31
Perfect score:   4573
Sequence:        1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2463

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters:      1027090

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155137_15173/app_query.fasta_1.7893
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_101_@runat_31052005_155137_15173 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
                 2:  /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
                 3:  /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
                 4:  /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
                 5:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                 6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3497	76.5	1005	2	US-07-743-357-1	Sequence 1, Appli
2	3493.5	76.4	1003	2	US-07-743-357-9	Sequence 9, Appli
3	3475	76.0	1015	3	US-08-463-210-9	Sequence 9, Appli
4	3475	76.0	1015	3	US-09-124-900-3	Sequence 3, Appli
5	3475	76.0	1015	4	US-08-463-028-9	Sequence 9, Appli
6	3475	76.0	1016	2	US-07-743-357-2	Sequence 2, Appli
7	3472	75.9	1003	4	US-09-309-572-17	Sequence 17, Appl
8	3472	75.9	1003	4	US-09-718-096-17	Sequence 17, Appl
9	3469	75.9	1016	2	US-07-743-357-5	Sequence 5, Appli
10	3468	75.8	1016	2	US-07-743-357-3	Sequence 3, Appli
11	3465	75.8	1016	2	US-07-743-357-4	Sequence 4, Appli
12	3459.5	75.7	1003	2	US-07-743-357-10	Sequence 10, Appl
13	3456.5	75.6	1003	2	US-07-743-357-8	Sequence 8, Appli
14	3449	75.4	1004	2	US-07-743-357-7	Sequence 7, Appli
15	3183.5	69.6	1014	4	US-09-319-588C-6	Sequence 6, Appli
16	3141	68.7	913	2	US-07-743-357-22	Sequence 22, Appl
17	3119.5	68.2	1350	4	US-09-952-060-35	Sequence 35, Appl
18	3078	67.3	913	2	US-07-743-357-6	Sequence 6, Appli
19	2822.5	61.7	875	4	US-09-952-060-6	Sequence 6, Appli
20	2817	61.6	850	4	US-09-952-060-2	Sequence 2, Appli
21	2798.5	61.2	875	4	US-09-952-060-8	Sequence 8, Appli
22	2793	61.1	850	4	US-09-952-060-4	Sequence 4, Appli
23	2789	61.0	562	3	US-09-117-217-14	Sequence 14, Appl
24	2789	61.0	562	4	US-09-735-487-14	Sequence 14, Appl
25	2758	60.3	560	4	US-10-205-641-1	Sequence 1, Appli
26	2507.5	54.8	1018	4	US-09-206-551-46	Sequence 46, Appl
27	2483.5	54.3	1016	4	US-09-625-972-24	Sequence 24, Appl
28	2259	49.4	441	4	US-09-991-258-16	Sequence 16, Appl
29	2212	48.4	1055	2	US-08-659-251-5	Sequence 5, Appli
30	2212	48.4	1055	3	US-09-256-490-5	Sequence 5, Appli
31	2212	48.4	1055	5	PCT-US96-11445-5	Sequence 5, Appli
32	2193.5	48.0	3080	6	5223423-4	Patent No. 5223423
33	2193.5	48.0	3080	6	5223423-4	Patent No. 5223423
34	2154	47.1	427	3	US-09-690-265-1	Sequence 1, Appli
35	1572	34.4	327	3	US-08-679-493A-72	Sequence 72, Appl
36	1555	34.0	327	3	US-08-679-493A-71	Sequence 71, Appl
37	1552	33.9	314	1	US-08-589-446-6	Sequence 6, Appli
38	1552	33.9	314	1	US-08-444-882-6	Sequence 6, Appli
39	1552	33.9	314	2	US-08-389-459A-6	Sequence 6, Appli
40	1552	33.9	314	3	US-08-987-867A-6	Sequence 6, Appli
41	1311	28.7	1150	3	US-09-238-303-9	Sequence 9, Appli
42	1311	28.7	1150	4	US-09-946-239-9	Sequence 9, Appli
43	1309	28.6	275	6	5252477-3	Patent No. 5252477
44	1309	28.6	275	6	5252477-3	Patent No. 5252477
45	1297	28.4	261	6	5320958-3	Patent No. 5320958

ALIGNMENTS

```
RESULT 1
US-07-743-357-1
; Sequence 1, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  1:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1005 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  protein
;     HYPOTHETICAL:  NO
;     FRAGMENT TYPE:  internal
;     ORIGINAL SOURCE:
;       ORGANISM:  Human immunodeficiency virus type 1
;       STRAIN:  HXB2
US-07-743-357-1

Alignment Scores:
Pred. No.:               1.22e-208     Length:         1005
Score:                   3497.00       Matches:        655
Percent Similarity:      95.56%        Conservative:   34
Best Local Similarity:   90.85%        Mismatches:     26
Query Match:             76.47%        Indels:         6
DB:                      2             Gaps:           3

US-09-610-313B-31 (1-2463) x US-07-743-357-1 (1-1005)

Qy     220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db       2 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 21

Qy     280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db      22 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 41

Qy     334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db      42 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 61

Qy     388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db      62 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 81

Qy     448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db      82 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 101

Qy     508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db     102 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 121

Qy     568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db     122 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 141

Qy     628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db     142 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 161

Qy     688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db     162 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 181

Qy     748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db     182 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 201

Qy     808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db     202 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 221

Qy     868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db     222 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 241

Qy     928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db     242 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 261

Qy     988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db     262 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 281

Qy    1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db     282 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 301

Qy    1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db     302 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 321

Qy    1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db     322 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 341

Qy    1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db     342 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 361

Qy    1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db     362 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 381

Qy    1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
Db     382 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 401

Qy    1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
Db     402 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 421

Qy    1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
Db     422 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 441

Qy    1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
Db     442 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 461

Qy    1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
Db     462 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 481

Qy    1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
Db     482 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 501

Qy    1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
Db     502 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 521

Qy    1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
Db     522 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 541

Qy    1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
Db     542 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 561

Qy    1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
Db     562 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 581

Qy    1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
Db     582 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 601

Qy    2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
Db     602 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 621
```

```
Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::|||:::|||||:::||||||||||||||||||||||||||||||   ||||||
Db        622 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 641

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              ||||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db        642 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 661

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        662 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 681

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        682 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 701

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        702 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 721

Qy       2362 GAT 2364
              |||
Db        722 Asp 722
```

```
RESULT 2
US-07-743-357-9
; Sequence 9, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1003 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  MAL
US-07-743-357-9
```

```
Alignment Scores:
Pred. No.:               2.01e-208      Length:        1003
Score:                   3493.50        Matches:       652
Percent Similarity:      95.97%         Conservative:  39
Best Local Similarity:   90.56%         Mismatches:    24
Query Match:             76.39%         Indels:        5
DB:                      2              Gaps:          3

US-09-610-313B-31 (1-2463) x US-07-743-357-9 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||:|:::|||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
                 ||||||||||||||||||||||||||||:::|||         ||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACCCTG 390
              ||||||    ||||||||||||||||         :::     :::||||||||||||||||
Db         41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPheSerPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||||:::|||:::::::::||||||||||||:::|||||||||||||||||||
Db         61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              |||||||||||||||||||||||||||:::::::||||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||||||||||||||||||||:::|||:::||||||||||||||||||||||||
Db        121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              |||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              ||||||||||||||||||||||||||     ||||||::::::||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              |||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
```

```
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
              ||||||||||||||||     :::||||||||||||||||||||||
Db        321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy       1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
              ||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy       1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
              :::|||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy       1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400

Qy       1405 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1464
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy       1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1524
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440

Qy       1525 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1584
              |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460

Qy       1585 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1644
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1645 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1704
              ||||||||||||||||||||      ||||||||||||||||||||||||||     :::
Db        481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500

Qy       1705 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1764
              |||||||||||||||||||||||||||:::::::::::||||||||||||||||||||||
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520

Qy       1765 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1824
              |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db        521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540

Qy       1825 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884
              ||||||||||||||||||||||||||||||||||||||||   |||||||||:::|||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy       1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1945 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2004
              ||||||||||||   |||||||||:::|||||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       2005 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2064
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||:::
Db        601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2065 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2124
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy       2125 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2184
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2185 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2244
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

Qy       2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2304
              :::|||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2305 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              |||:::|||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

```
RESULT 3
US-08-463-210-9
; Sequence 9, Application US/08463210
; Patent No. 6001977
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, Nancy T.
;    APPLICANT:  GALLO, Robert C.
;    APPLICANT:  WONG-STAAL, Flossie
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF HTLV-III DNA
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morgan & Finnegan, L.L.P.
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10154-0053
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,210
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/693,866
;      FILING DATE:  23-JAN-1985
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/659,339
;      FILING DATE:  10-OCT-1984
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Serunian, Leslie A.
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4193US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1015 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HTLV-III
;    FEATURE:
;      NAME/KEY:  Protein
;      LOCATION:  1..1015
;      OTHER INFORMATION:  /note= "pol protein of HTLV-III"
US-08-463-210-9

Alignment Scores:
Pred. No.:                2.82e-207       Length:         1015
Score:                    3475.00         Matches:        654
Percent Similarity:       93.72%          Conservative:   33
Best Local Similarity:    89.22%          Mismatches:     28
Query Match:              75.99%          Indels:         18
DB:                       3               Gaps:           4

US-09-610-313B-31 (1-2463) x US-08-463-210-9 (1-1015)
```

```
Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy    268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db     21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy    304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db     41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy    358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db     61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy    412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db     81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy    472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db    101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy    532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db    121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy    592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db    141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy    652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db    161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy    712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db    181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy    772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db    201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy    832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db    221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy    892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db    241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy    952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db    261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy   1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db    281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy   1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db    301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy   1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db    321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy   1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db    341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy   1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db    361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy   1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
Db    381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy   1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
Db    401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy   1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
Db    421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy   1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
Db    441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy   1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
Db    461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy   1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
Db    481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy   1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
Db    501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy   1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
Db    521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy   1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
Db    541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy   1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
Db    561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy   1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
Db    581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy   1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
Db    601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy   2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
Db    621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy   2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
Db    641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy   2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
Db    661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy   2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
Db    681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy   2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
Db    701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy   2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
```

```
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 4
US-09-124-900-3
; Sequence 3, Application US/09124900
; Patent No. 6268484
; GENERAL INFORMATION:
;  APPLICANT:  KATINGER, Hermann
;  APPLICANT:  BUCHACHER, Andrea
;  APPLICANT:  ERNST, Wolfgang
;  APPLICANT:  BALLAUN, Claudia
;  APPLICANT:  PURTSCHER, Martin
;  APPLICANT:  TRKOLA, Alexandra
;  APPLICANT:  PREDL, Renate
;  APPLICANT:  SCHMATZ, Christine
;  APPLICANT.  KLIMA, Annelies
;  APPLICANT:  STEINDL, Franz
;  APPLICANT:  MUSTER, Thomas
;  TITLE OF INVENTION:   HIV-Vaccines
;  FILE REFERENCE:   1939-112P
;  CURRENT APPLICATION NUMBER: US/09/124,900
;  CURRENT FILING DATE:  1998-07-30
;  PRIOR APPLICATION NUMBER:    PCT/EP95/01481
;  PRIOR FILING DATE:   1995-04-19
;  NUMBER OF SEQ ID NOS:   11
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 1015
;   TYPE: PRT
;   ORGANISM: Human imunodeficiency virus type 1
US-09-124-900-3

Alignment Scores:
Pred. No.:               2.82e-207      Length:         1015
Score:                   3475.00        Matches:        654
Percent Similarity:      93.72%         Conservative:   33
Best Local Similarity:   89.22%         Mismatches:     28
Query Match:             75.99%         Indels:         18
DB:                      3              Gaps:           4

US-09-610-313B-31 (1-2463) x US-09-124-900-3 (1-1015)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
```

```
Qy       1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
              |||||:::|||||   ||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
              ||||||        ||||||||||||||||||||||||||||||||:::||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
              |||:::|||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
              ||||||||||||||||::::::||||||||||||:::|||    ||||||    ||||||
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
              ||||||||||||||||||||||||||||    ||||||||||||||||||||    ||||||
Db        641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
              :::|||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
                    |||||||||||:::||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 5
US-08-463-028-9
; Sequence 9, Application US/08463028
; Patent No. 6610476
; GENERAL INFORMATION:
;   APPLICANT: CHANG, Nancy T.
;   APPLICANT: GALLO, Robert C.
;   APPLICANT: WONG-STAAL, Flossie
;   TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Morgan & Finnegan, L.L.P.
;   STREET: 345 Park Avenue
;   CITY: New York
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 10154-0053
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/463,028
;   FILING DATE: 05-JUN-1995
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 06/693,866
;   FILING DATE: 23-JAN-1985
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 06/659,339
;   FILING DATE: 10-OCT-1984
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Serunian, Leslie A.
;   REGISTRATION NUMBER: 35,353
;   REFERENCE/DOCKET NUMBER: 2026-4193US3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 758-4800
;   TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1015 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ORIGINAL SOURCE:
;   ORGANISM: HTLV-III
;   FEATURE:
;   NAME/KEY: Protein
;   LOCATION: 1..1015
;   OTHER INFORMATION: /note= "pol protein of HTLV-III"
US-08-463-028-9

```
Alignment Scores:
Pred. No.:              2.82e-207     Length:        1015
Score:                  3475.00       Matches:       654
Percent Similarity:     93.72%        Conservative:  33
Best Local Similarity:  89.22%        Mismatches:    28
Query Match:            75.99%        Indels:        18
DB:                     4             Gaps:          4
```

US-09-610-313B-31 (1-2463) x US-08-463-028-9 (1-1015)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC----------- 267
              |||||||||||||||||||||||    |||||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
                                      |||||||||    |||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||     :::|||   ||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      |||||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::||||||||:::||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              ||||||||||||||||||||||||||||||||||||||||||||||||    ||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              |||||||||:::|||||||||||||||||||||||||||||||||||:::||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
```

```
Qy    712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db    181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy    772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db    201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy    832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db    221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy    892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db    241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy    952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db    261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy   1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db    281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy   1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db    301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy   1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db    321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy   1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db    341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy   1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db    361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy   1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
Db    381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy   1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
Db    401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy   1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
Db    421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy   1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
Db    441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy   1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
Db    461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy   1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
Db    481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy   1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
Db    501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy   1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
Db    521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy   1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
Db    541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy   1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
Db    561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy   1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
Db    581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy   1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
Db    601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy   2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
Db    621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy   2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
Db    641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy   2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
Db    661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy   2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
Db    681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy   2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
Db    701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy   2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db    721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

```
RESULT 6
US-07-743-357-2
; Sequence 2, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
```

```
;     TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1016 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  BH102
US-07-743-357-2

Alignment Scores:
Pred. No.:               2.82e-207      Length:         1016
Score:                   3475.00        Matches:        654
Percent Similarity:      93.72%         Conservative:   33
Best Local Similarity:   89.22%         Mismatches:     28
Query Match:             75.99%         Indels:         18
DB:                      2              Gaps:           4

US-09-610-313B-31 (1-2463) x US-07-743-357-2 (1-1016)

Qy      220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db        1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy      268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db       21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy      304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db       41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy      358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db       61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy      412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db       81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy      472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db      101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy      532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db      121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy      592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db      141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy      652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db      161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy      712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db      181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy      772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db      201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy      832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db      221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy      892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db      241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy      952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db      261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy     1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db      281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy     1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db      301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy     1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db      321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy     1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db      341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy     1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db      361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy     1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
Db      381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy     1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
Db      401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy     1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
Db      421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy     1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
Db      441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy     1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
Db      461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy     1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
Db      481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy     1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
Db      501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy     1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
Db      521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy     1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
Db      541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy     1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
Db      561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy     1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
Db      581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy     1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
```

```
                  |||:::||||||||||||||||||||||||||||||||||||||||||||:::|||
Db          601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy         2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
                ||||||||||||||||:::::|||||||||||:::|||   ||||||   ||||||
Db          621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy         2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
                |||||||||||||||||||||||||   |||||||||||||||||||   ||||||
Db          641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy         2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy         2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy         2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
                :::|||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db          701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy         2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
                   ||||||||||:::|||||||||||||||||||
Db          721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 7
US-09-309-572-17
; Sequence 17, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
;  APPLICANT: Heinrich-Pette-Institut
;  TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
;  FILE REFERENCE: P50489
;  CURRENT APPLICATION NUMBER: US/09/309,572
;  CURRENT FILING DATE: 1999-05-11
;  EARLIER APPLICATION NUMBER: DE 198 56 463
;  EARLIER FILING DATE: 1998-11-26
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
;   FEATURE:
;   OTHER INFORMATION: pol polyprotein
US-09-309-572-17

Alignment Scores:
Pred. No.:               4.32e-207      Length:        1003
Score:                   3472.00        Matches:       652
Percent Similarity:      95.15%         Conservative:  34
Best Local Similarity:   90.43%         Mismatches:    29
Query Match:             75.92%         Indels:        6
DB:                      4              Gaps:          3

US-09-610-313B-31 (1-2463) x US-09-309-572-17 (1-1003)

Qy          220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
                ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db            1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy          280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                   |||||||||||||||||   |||||||||||||||      |||   :::|||
Db           21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy          334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                   ||||||||||||||:::||||||||||||:::      :::|||||||||||||||
Db           41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy          388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
                ||||||||||||||||||||:::||||||:::|||||||||:::||||||||||||||||
Db           61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy          448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
                ||||||||||||||||||||||||||||||||||:::|||||||||:::||||||||||
Db           81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy          508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
                |||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db          101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120

Qy          568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
                ||||||||||||   ||||||||||||||||||:::||||||||||||||||||||||||
Db          121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy          628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
                ||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||
Db          141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy          688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy          748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
                ||||||||||||||||||||||||   ||||||   ||||||||||||||||||
Db          181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy          808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
                ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db          201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy          868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy          928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
                ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db          241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSer 260

Qy          988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
                ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db          261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy         1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy         1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
                ||||||||||||||||||||||||||||||||:::|||||||||   |||||||||
Db          301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320

Qy         1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
                ||||||||||||||||||||   :::||||||:::||||||||||||||||
Db          321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy         1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                   |||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db          341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy         1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
                |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy         1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy         1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
                ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db          401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy         1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
                ||||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db          421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
```

```
Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                |||||||||||::::::||||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              :::||||||||||||||||||||         ||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              |||||||||||||||||||||||||||||||:::|||:::   |||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
              ||||||||||||:::||||||||||||||||||||||||||   |||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::|||   ||||||:::||||||||||||||||||||||||||||||   ||||||
Db        621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              |||||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::|||   |||||||||   |||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              |||
Db        721 Asp 721
```

RESULT 8
US-09-718-096-17
; Sequence 17, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718,096
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: DE 19856463.5
; PRIOR FILING DATE: 1998-11-26
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
;   FEATURE:
;   OTHER INFORMATION: pol polyprotein
US-09-718-096-17

Alignment Scores:			
Pred. No.:	4.32e-207	Length:	1003
Score:	3472.00	Matches:	652
Percent Similarity:	95.15%	Conservative:	34
Best Local Similarity:	90.43%	Mismatches:	29
Query Match:	75.92%	Indels:	6
DB:	4	Gaps:	3

US-09-610-313B-31 (1-2463) x US-09-718-096-17 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                 ||||||||||||||||   ||||||||||||||         |||   :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||:::|||||||||||:::         :::||||||||||||||
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||||:::|||||||:::|||||||||:::||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              |||||||||||||||||||||||||||||||||||:::|||||||||:::||||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              ||||||||||||   |||||||||||||||||||||:::||||||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              |||||||||:::|||||||||:::||||||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||      ||||||    ||||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||||:::|||||||||   ||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              ||||||||||||||||||     :::||||||:::||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 |||||||||||||||||||||||||||||||||||||||     ||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                 |||||||||||:::::::||||||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              ||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              :::||||||||||||||||||||||       ||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              ||||||||||||||||||||||||||||||||:::|||:::   ||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
              ||||||||||||:::||||||||||||||||||||||||||||   ||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::|||   ||||||:::||||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::|||   |||||||||   ||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              |||
Db        721 Asp 721
```

RESULT 9
US-07-743-357-5
; Sequence 5, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
;   APPLICANT: Kang, Yong C.
;   TITLE OF INVENTION: Polypeptide having immunological
;   TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: KIRBY EADES GALE BAKER
;   STREET: Box 3432, Station D
;   CITY: Ottawa
;   STATE: Ontario
;   COUNTRY: Canada
;   ZIP: K1M 1H8
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/743,357
;   FILING DATE: 21-AUG-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/CA90/00062
;   FILING DATE: 23-FEB-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Gale, Edwin J.
;   REGISTRATION NUMBER: 28,584
;   REFERENCE/DOCKET NUMBER: 30924-2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (613) 237-6900
;   TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1016 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   FRAGMENT TYPE: internal
;   ORIGINAL SOURCE:
;   ORGANISM: Human immunodeficiency virus type 1
;   STRAIN: BRU

```
US-07-743-357-5

Alignment Scores:
Pred. No.:                6.64e-207     Length:          1016
Score:                    3469.00       Matches:         652
Percent Similarity:       93.86%        Conservative:    36
Best Local Similarity:    88.95%        Mismatches:      27
Query Match:              75.86%        Indels:          18
DB:                       2             Gaps:            4

US-09-610-313B-31 (1-2463) x US-07-743-357-5 (1-1016)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy    268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db     21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy    304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db     41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArg 60

Qy    358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db     61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy    412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db     81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy    472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db    101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy    532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db    121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy    592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db    141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy    652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db    161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy    712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db    181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy    772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db    201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy    832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db    221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy    892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db    241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy    952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db    261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy   1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db    281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy   1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db    301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy   1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db    321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340

Qy   1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db    341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy   1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db    361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy   1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
Db    381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy   1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
Db    401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy   1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
Db    421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy   1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
Db    441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy   1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
Db    461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy   1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
Db    481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy   1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
Db    501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy   1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
Db    521 TyrAlaArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy   1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
Db    541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy   1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
Db    561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy   1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
Db    581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy   1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
Db    601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaSerArgGluThrLysLeuGly 620

Qy   2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
Db    621 LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValValThrLeuThrAspThrThr 640

Qy   2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
Db    641 AsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy   2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
```

```
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
              :::|||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              ||||||||||||:::||||||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 10
US-07-743-357-3
; Sequence 3, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1016 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  BH5
US-07-743-357-3

Alignment Scores:
Pred. No.:              7.66e-207      Length:         1016
Score:                  3468.00        Matches:        652
Percent Similarity:     93.72%         Conservative:   35
Best Local Similarity:  88.95%         Mismatches:     28
Query Match:            75.84%         Indels:         18
DB:                     2              Gaps:           4

US-09-610-313B-31 (1-2463) x US-07-743-357-3 (1-1016)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC----------- 267
              ||||||||||||||||||||||||   ||||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 -----------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
                                     |||||||||   ||||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||   :::|||      ||||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      ||||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              |||||||||||||||||||||||||||||||||||||||||:||||||   |||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::||||||||||||||||||||||||||||||||:::||||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
              |||      ||||||   |||||||||||||||||||||||||:::||||||||||||||
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        241 LeuValAspPheArgGluLeuAsnArgArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
              ||||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        301 SerIleAsnAsnGluThrProGlySerGlyTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
              :::||||||:::|||||||||||||||      |||||||||||||||||||||||||||
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
```

```
                 ||||||||||||||||||    ||||||||||||||||:::||||||||||||||||
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 PheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
              ||||||||||||||||||:::|||||||||   ||||||||||||:::||||||||
Db        401 LeuHisProAspLysTrpThrIleGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
              ||||||||||||||||||||||||||||||||||||   |||||||||||:::::::::
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
              ||||||:::||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
              ||||||     |||||||||||||||||||||||||||||||||||:::|||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
              |||:::|||||||||||||||||||||||||||||:::||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaSerArgGluThrLysLeuGly 620

Qy       2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
              |||||||||||||||||:::||||||||||||||||:::|||:::||||||    ||||||
Db        621 LysAlaGlyTyrValThrAsnArgGlyArgGlnLysValValThrLeuThrHisThrThr 640

Qy       2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||  ||||||
Db        641 AsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
              :::||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
                 ||||||||||:::|||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 11

```
US-07-743-357-4
; Sequence 4, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1016 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  PV22
US-07-743-357-4

Alignment Scores:
Pred. No.:              1.18e-206     Length:         1016
Score:                  3465.00       Matches:        651
Percent Similarity:     93.72%        Conservative:   36
Best Local Similarity:  88.81%        Mismatches:     28
Query Match:            75.77%        Indels:         18
DB:                     2             Gaps:           4

US-09-610-313B-31 (1-2463) x US-07-743-357-4 (1-1016)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC----------- 267
              |||||||||||||||||||||||    |||||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
                                      |||||||||    ||||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              ||||||||||||||||       |||    :::|||    ||||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
```

```
Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      |||||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::||||||||||||||||||||||||||||||||||||:::||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
              |||      ||||||   ||||||||||||||||||||||||:::|||||||||||||||
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||   
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
              :::||||||:::|||||||||||||||            |||||||||||||||||||||
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
              |||||||||||||||||||||||||||||||||   ||||||||||||:::|||||||||
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
              ||||||||||||||||||||||||||||||||||||   ||||||||||||:::::::::
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
              ||||||:::||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
              ||||||      ||||||||||||||||||||||||||||||||||||:::|||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
              |||:::|||||||||||||||||||||||||||||||||||||||||||||::::::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGly 620

Qy       2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
              ||||||||||||:::|||::::::||||||||||||:::|||   ||||||   ||||||
Db        621 LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
              |||||||||||||||||||||||||||   |||||||||||||||||||||   ||||||
Db        641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSer 680

Qy       2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlnLysValTyrLeu 700

Qy       2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
              :::||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
                 ||||||||||||:::||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 12
US-07-743-357-10
; Sequence 10, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
;   APPLICANT: Kang, Yong C.
;   TITLE OF INVENTION: Polypeptide having immunological
;   TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:

```
;       ADDRESSEE:  KIRBY EADES GALE BAKER
;       STREET:  Box 3432, Station D
;       CITY:  Ottawa
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  K1M 1H8
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/743,357
;       FILING DATE:  21-AUG-1991
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/CA90/00062
;       FILING DATE:  23-FEB-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Gale, Edwin J.
;       REGISTRATION NUMBER:  28,584
;       REFERENCE/DOCKET NUMBER:  30924-2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (613) 237-6900
;       TELEFAX:  (613) 237-0045
;   INFORMATION FOR SEQ ID NO:  10:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1003 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  protein
;     HYPOTHETICAL:  NO
;     FRAGMENT TYPE:  internal
;     ORIGINAL SOURCE:
;       ORGANISM:  Human immunodeficiency virus type 1
;       STRAIN:  ELI
US-07-743-357-10

Alignment Scores:
Pred. No.:              2.56e-206       Length:         1003
Score:                  3459.50         Matches:        646
Percent Similarity:     95.28%          Conservative:   40
Best Local Similarity:  89.72%          Mismatches:     29
Query Match:            75.65%          Indels:         5
DB:                     2               Gaps:           3

US-09-610-313B-31 (1-2463) x US-07-743-357-10 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||:::|||||||||||||||||||||||     |||         :::|||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerProLysGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
                   |||||||||||||||||||||||||||:::|||   |||   ||||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTG 390
              |||:::    ||||||||||||||||||||:::        ||||||||||||||||||||
Db         41 SerLysThrGlyAlaGluArgGlnGlyThrValSerPheAsnPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||||||||:::||||||:::||||||||:::|||||||||||||||||||
Db         61 TrpGlnArgProLeuValAlaIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||:::|||||||||||||||||||||||:::||||||||||||||||||:::
Db        121 IleCysGlyGlnLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              ||||||||:::|||||||||:::||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              |||||||||||||||||||||||||||      ||||||    :::||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrAspMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||::::::|||||||||||||||||||||||||||||||||:::||||||||||||||||
Db        201 IleSerArgIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleSerSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
              ||||||||||||||||||      :::|||||||||:::||||||||||||
Db        321 IleLeuGluProPheArgLysGlnAsnProGluMetValIleTyrGlnTyrMetAspAsp 340

Qy       1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
              |||||||||||||||||||||||||||||||||||||      |||||||||:::||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLysLeuArg 360

Qy       1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
              :::||||||||||||||||||||||||||     ||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuArgTrpGlyPheThrArgProAspLysLysHisGlnLysGluProPro 380

Qy       1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404
              ||||||||||||||||||||||||||||||||||||||||||||||||    |||:::|||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400

Qy       1405 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1464
              |||||||||||||||||||||||||||||||||     ||||||     :::|||||||||
Db        401 ProGluLysGluSerTrpThrValAsnAspIleGlnAsnLeuValGluArgLeuAsnTrp 420

Qy       1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr 440

Qy       1525 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1584
              |||||||||||::::::::||||||||||||||||||||||||||||||||||||||||
Db        441 LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460

Qy       1585 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1644
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1645 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1704
```

```
                  ||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy       1705 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1764
              |||||||||||||||||||||||||:::||||||   |||||||||||||||||||
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520

Qy       1765 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1824
              |||||  ||||||||||:::|||:::  |||||||||||||||||||:::|||
Db        521 GlnLeuAlaGluAlaValGlnArgIleSerThrGluSerIleValIleTrpGlyArgThr 540

Qy       1825 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884
              ||||||||||||||||||||||||||||||||||||||||||||||||   :::||||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpAlaGluTyrTrp 560

Qy       1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1945 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       2005 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2064
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||:::
Db        601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2065 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2124
              |||  ||||||:::||||||||||||||||||||||||||||||||   ||||||||
Db        621 ValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640

Qy       2125 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2184
              |||||||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2185 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIle 680

Qy       2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2304
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2305 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              |||:::|||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 13

```
US-07-743-357-8
; Sequence 8, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1003 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  RF
US-07-743-357-8

Alignment Scores:
Pred. No.:              3.93e-206      Length:        1003
Score:                  3456.50        Matches:       647
Percent Similarity:     95.28%         Conservative:  39
Best Local Similarity:  89.86%         Mismatches:    29
Query Match:            75.58%         Indels:        5
DB:                     2              Gaps:          3

US-09-610-313B-31 (1-2463) x US-07-743-357-8 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||:::|||||||||||||||||||||||||||||||||      ||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluLeuSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
                 ||||||||||||||||  |||||||||||||   |||   |||||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSerLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTG 390
              |||||||||||   :::|||||||||||:::      :::||||||||||||||||
Db         41 SerGluAlaGlyGluAspArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||:::|||::::::|||:::|||||||||:::||||||||||||||||
Db         61 TrpGlnArgProIleValThrValLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              ||||||||   ||||||||||||||||||:::||||||||||||||||||||||||
Db        121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              ||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              ||||||||||||||||||||||||      ||||||   ||||||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLys 200
```

```
Qy      811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
            |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      201 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy      871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy      931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy      991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
            |||||||||||||||||||||||||||||||||||||||||||||::::::|||||||||
Db      261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysGluPheArgLys 280

Qy     1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
            |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db      281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProArgIleArgTyrGlnTyr 300

Qy     1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
            |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db      301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy     1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
            |||||||||||||||:::   :::||||||||||||||||||||||||
Db      321 IleLeuGluProPheLysLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy     1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
            |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db      341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgIleLysIleGluGluLeuArg 360

Qy     1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
            :::|||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db      361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy     1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db      381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleValLeu 400

Qy     1405 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1464
            |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db      401 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy     1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1524
            |||||||||||||||   ||||||||||||:::||||||||||||||||||||||||
Db      421 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 440

Qy     1525 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1584
            ||||||||||||::::::|||   |||:::||||||||||||||||||||||||||||||
Db      441 LysAlaLeuThrGluValValGlnLeuThrLysGluAlaGluLeuGluLeuAlaGluAsn 460

Qy     1585 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1644
            |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||:::
Db      461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy     1645 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1704
            ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db      481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy     1705 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1764
            ||||||||||||||||||||||||:::||||||   ||||||||||||||||||||||||
Db      501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520

Qy     1765 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1824
            |||||||||||||||||||||:::|||   ||||||||||||||||||||||||||||||
Db      521 GlnLeuThrGluAlaValGlnLysValAlaThrGluSerIleValIleTrpGlyLysThr 540

Qy     1825 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884
            |||||||||:::|||||||||||||||||||||||||||   |||||||||:::||||||
Db      541 ProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy     1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1944
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy     1945 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2004
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy     2005 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2064
            ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||:::
Db      601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy     2065 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2124
            |||||||||:::||||||||||||||||||||||||||||||||||||   |||||||||
Db      621 ValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy     2125 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2184
            ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db      641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy     2185 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2244
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680

Qy     2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2304
            |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db      681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy     2305 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
            |||:::|||:::|||||||||   ||||||||||||||||||||||||||||||||||||
Db      701 GlnValAspArgLeuValSerThrGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 14
US-07-743-357-7
; Sequence 7, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1004 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein

```
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  SF2
US-07-743-357-7

Alignment Scores:
Pred. No.:              1.15e-205    Length:          1004
Score:                  3449.00      Matches:         646
Percent Similarity:     94.87%       Conservative:    38
Best Local Similarity:  89.60%       Mismatches:      31
Query Match:            75.42%       Indels:          6
DB:                     2            Gaps:            3

US-09-610-313B-31 (1-2463) x US-07-743-357-7 (1-1004)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy    280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------GACAACCCC 333
Db     21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40

Qy    334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db     41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy    388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db     61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy    448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db     81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100

Qy    508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db    101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProVal 120

Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db    321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy   1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db    341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy   1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db    361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy   1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
Db    381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMet 400

Qy   1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
Db    401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy   1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
Db    421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy   1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
Db    441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy   1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
Db    461 AsnArgGluIleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeu 480

Qy   1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
Db    481 ValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy   1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
Db    501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy   1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
Db    521 LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys 540

Qy   1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
Db    541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyr 560

Qy   1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
Db    561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy   1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
Db    581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy   2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
Db    601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy   2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
Db    621 ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy   2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
```

```
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720

Qy       2362 GAT 2364
Db        721 Asp 721
```

RESULT 15
US-09-319-588C-6
; Sequence 6, Application US/09319588C
; Patent No. 6509018
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAUCLERE, Philippe
; APPLICANT: LOUSSERT-AJAKA, Ibtissam
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSSI, Francoise
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/09/319,588C
; CURRENT FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1014
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-09-319-588C-6

Alignment Scores:

Pred. No.:	3.17e-189	Length:	1014
Score:	3183.50	Matches:	583
Percent Similarity:	90.25%	Conservative:	74
Best Local Similarity:	80.08%	Mismatches:	58
Query Match:	69.62%	Indels:	13
DB:	4	Gaps:	4

US-09-610-313B-31 (1-2463) x US-09-319-588C-6 (1-1014)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProProAspAsn 20

Qy        280 AAC------CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG--------------- 318
Db         21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40

Qy        319 ---CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG---------GGCACC 366
Db         41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db         61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db         81 GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluGluLeuGln 100

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db        101 LeuGluGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArg 120

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db        121 GlnTyrAspAsnIleThrValAspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db        141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db        161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db        181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db        201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db        221 ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db        241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db        261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db        281 ProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGlu 300

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db        301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db        321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340

Qy       1207 GTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db        341 IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db        361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db        381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db        401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db        421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
```

```
Db        441 LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||:::||||||:::||||||||||||
Db        461 AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              |||||||||   |||:::|||||||||||||||||||||||||||      |||||||||
Db        481 TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThr 500

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        501 TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              :::||||||||||||||:::|||||||||   |||   |||:::|||:::|||   |||
Db        521 SerAlaHisThrAsnAspIleLysGlnLeuValGluValValArgLysValAlaThrGlu 540

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              |||||||||||||||||||||||||||||||||||||||:::|||||||||   |||
Db        541 SerIleValIleTrpGlyLysThrProLysPheArgLeuProValGlnLysGluValTrp 560

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              |||   ||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db        561 GluAlaTrpTrpThrAspHisTrpGlnAlaThrTrpIleProGluTrpGluPheValAsn 580

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||   |||||||||   |||||||||
Db        581 ThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGluProIleSerGlyAlaGlu 600

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||:::||||||||||||:::|||
Db        601 ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              |||||||||||||||||||:::||||||:::   :::|||||||||||||||   |||
Db        621 ThrAspArgGlyArgGlnLysValValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||   :::|||||||||:::||||||   :::||||||||||||||||||
Db        641 LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||:::|||||||||   :::||||||||||||||||||||||||||||||
Db        661 SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              :::|||||||||||||:::|||||||||||||||:::|||||||||||||||||||||||
Db        681 SerGlnIleIleGluGluLeuIleLysLysGluArgValTyrLeuSerTrpValProAla 700

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              |||||||||||||||||||||||||:::||||||||||||||||   ||||||||||||
Db        701 HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720

Qy       2341 GTGCTGTTCCTGGACGGCATCGAT 2364
              :::||||||||||||||||||:::
Db        721 IleLeuPheLeuAspGlyIleGlu 728
```

Search completed: June 2, 2005, 05:12:37
Job time : 109 secs

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  2, 2005, 03:13:23 ; Search time 33.9172 Seconds
                                          (without alignments)
                                          10815.332 Million cell updates/sec

Title:          US-09-610-313B-32
Perfect score:  4555
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2457

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters:      1027090

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155137_15173/app_query.fasta_1.7893
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_101_@runat_31052005_155137_15173 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_AA:*
                1:  /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
                2:  /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
                3:  /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
                4:  /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
                5:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3454	75.8	1005	2	US-07-743-357-1	Sequence 1, Appli
2	3450.5	75.8	1003	2	US-07-743-357-9	Sequence 9, Appli
3	3432	75.3	1015	3	US-08-463-210-9	Sequence 9, Appli
4	3432	75.3	1015	3	US-09-124-900-3	Sequence 3, Appli
5	3432	75.3	1015	4	US-08-463-028-9	Sequence 9, Appli
6	3432	75.3	1016	2	US-07-743-357-2	Sequence 2, Appli
7	3429	75.3	1003	4	US-09-309-572-17	Sequence 17, Appl
8	3429	75.3	1003	4	US-09-718-096-17	Sequence 17, Appl
9	3426	75.2	1016	2	US-07-743-357-5	Sequence 5, Appli
10	3425	75.2	1016	2	US-07-743-357-3	Sequence 3, Appli
11	3422	75.1	1016	2	US-07-743-357-4	Sequence 4, Appli
12	3416.5	75.0	1003	2	US-07-743-357-10	Sequence 10, Appl
13	3413.5	74.9	1003	2	US-07-743-357-8	Sequence 8, Appli
14	3406	74.8	1004	2	US-07-743-357-7	Sequence 7, Appli
15	3140.5	68.9	1014	4	US-09-319-588C-6	Sequence 6, Appli
16	3098	68.0	913	2	US-07-743-357-22	Sequence 22, Appl
17	3076.5	67.5	1350	4	US-09-952-060-35	Sequence 35, Appl
18	3035	66.6	913	2	US-07-743-357-6	Sequence 6, Appli
19	2779.5	61.0	875	4	US-09-952-060-6	Sequence 6, Appli
20	2774	60.9	850	4	US-09-952-060-2	Sequence 2, Appli
21	2755.5	60.5	875	4	US-09-952-060-8	Sequence 8, Appli
22	2750	60.4	850	4	US-09-952-060-4	Sequence 4, Appli
23	2746	60.3	562	3	US-09-117-217-14	Sequence 14, Appl
24	2746	60.3	562	4	US-09-735-487-14	Sequence 14, Appl
25	2715	59.6	560	4	US-10-205-641-1	Sequence 1, Appli
26	2464.5	54.1	1018	4	US-09-206-551-46	Sequence 46, Appl
27	2445.5	53.7	1016	4	US-09-625-972-24	Sequence 24, Appl
28	2216	48.6	441	4	US-09-991-258-16	Sequence 16, Appl
29	2173	47.7	1055	2	US-08-659-251-5	Sequence 5, Appli
30	2173	47.7	1055	3	US-09-256-490-5	Sequence 5, Appli
31	2173	47.7	1055	5	PCT-US96-11445-5	Sequence 5, Appli
32	2154.5	47.3	3080	6	5223423-4	Patent No. 5223423
33	2154.5	47.3	3080	6	5223423-4	Patent No. 5223423
34	2111	46.3	427	3	US-09-690-265-1	Sequence 1, Appli
35	1529	33.6	327	3	US-08-679-493A-72	Sequence 72, Appl
36	1512	33.2	327	3	US-08-679-493A-71	Sequence 71, Appl
37	1509	33.1	314	1	US-08-589-446-6	Sequence 6, Appli
38	1509	33.1	314	1	US-08-444-882-6	Sequence 6, Appli
39	1509	33.1	314	2	US-08-389-459A-6	Sequence 6, Appli
40	1509	33.1	314	3	US-08-987-867A-6	Sequence 6, Appli
41	1309	28.7	275	6	5252477-3	Patent No. 5252477
42	1309	28.7	275	6	5252477-3	Patent No. 5252477
43	1272	27.9	1150	3	US-09-238-303-9	Sequence 9, Appli
44	1272	27.9	1150	4	US-09-946-239-9	Sequence 9, Appli
45	1254	27.5	261	6	5320958-3	Patent No. 5320958

ALIGNMENTS

```
RESULT 1
US-07-743-357-1
; Sequence 1, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  1:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1005 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  HXB2
US-07-743-357-1

Alignment Scores:
Pred. No.:                9.89e-208     Length:        1005
Score:                    3454.00       Matches:       651
Percent Similarity:       95.01%        Conservative:  34
Best Local Similarity:    90.29%        Mismatches:    28
Query Match:              75.83%        Indels:        8
DB:                       2             Gaps:          4

US-09-610-313B-32 (1-2457) x US-07-743-357-1 (1-1005)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db      2 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 21

Qy    280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db     22 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 41

Qy    334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db     42 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 61

Qy    388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db     62 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 81

Qy    448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db     82 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 101

Qy    508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db    102 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 121

Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    122 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 141

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    142 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 161

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    162 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 181

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    182 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 201

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    202 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 221

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    222 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 241

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    242 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 261

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    262 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 281

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    282 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 301

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    302 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 321

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db    322 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 341

Qy   1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db    342 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 361

Qy   1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db    362 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 381

Qy   1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
Db    382 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 401

Qy   1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db    402 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 421

Qy   1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db    422 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 441

Qy   1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db    442 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 461

Qy   1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db    462 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 481

Qy   1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db    482 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 501

Qy   1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db    502 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 521

Qy   1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db    522 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 541

Qy   1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db    542 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 561

Qy   1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db    562 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 581

Qy   1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db    582 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 601

Qy   1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db    602 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 621
```

```
Qy     2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
            :::|||:::|||||:::|||||||||||||||||||||||||||||||    ||||||
Db      622 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 641

Qy     2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
            ||||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db      642 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 661

Qy     2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
            ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db      662 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 681

Qy     2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
            ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db      682 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 701

Qy     2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
            ||||||:::|||||||||||||||   |||||||||||||||||||||||||||||||||
Db      702 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 721

Qy     2356 GAT 2358
            |||
Db      722 Asp 722
```

RESULT 2
US-07-743-357-9
; Sequence 9, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: MAL
US-07-743-357-9

```
Alignment Scores:
Pred. No.:              1.63e-207     Length:        1003
Score:                  3450.50       Matches:       648
Percent Similarity:     95.42%        Conservative:  39
Best Local Similarity:  90.00%        Mismatches:    26
Query Match:            75.75%        Indels:        7
DB:                     2             Gaps:          4

US-09-610-313B-32 (1-2457) x US-07-743-357-9 (1-1003)

Qy      220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
            ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy      280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
               |||||||||||||||||||||||||||||:::|||         ||||||
Db       21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40

Qy      337 AGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACCCTG 390
            ||||||   |||||||||||||||||          :::   :::||||||||||||||||
Db       41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPheSerPheProGlnIleThrLeu 60

Qy      391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
            ||||||||||||:::|||:::::::::||||||||||||||||||:::||||||||||||
Db       61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy      451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
            ||||||||||||||||||||||||||||||||::::::||||||||||||||||||||||
Db       81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy      511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy      571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
            ||||||||||||||||||||||||:::|||:::|||||||||||||||||||||||||||
Db      121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140

Qy      631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
            ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db      141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy      691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
            |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db      161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180

Qy      751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
            |||||||||||||||||||||||||      ||||||::::::|||||||||||||||||
Db      181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200

Qy      811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
            |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy      871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
            ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db      221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy      931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy      991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy     1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy     1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
            |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
```

```
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
              ||||||||||||||||    :::|||||||||||||||||||||
Db        321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy       1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy       1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
              :::|||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy       1345 TTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1398
              ||||||            ||||||||||||||||||||||||||||||||:::|||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400

Qy       1399 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1458
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy       1459 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1518
              ||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440

Qy       1519 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1578
              ||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460

Qy       1579 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1638
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1639 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1698
              ||||||||||||||||||||   |||||||||||||||||||||||||||   :::
Db        481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500

Qy       1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1758
              |||||||||||||||||||||||||||:::::::::::||||||||||||||||||
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520

Qy       1759 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1818
              ||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540

Qy       1819 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878
              ||||||||||||||||||||||||||||||||||||||   |||||||||:::|||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy       1879 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1938
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1939 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 1998
              ||||||||||||     |||||||:::||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       1999 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2058
              |.||||||||||||||    ||||||||||||||||||||||||||||||||||:::
Db        601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2059 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy       2119 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2178
              ||||||||||||    ||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2179 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2238
              ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

Qy       2239 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2298
              :::|||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2299 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||:::||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

```
RESULT 3
US-08-463-210-9
; Sequence 9, Application US/08463210
; Patent No. 6001977
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, Nancy T.
;    APPLICANT:  GALLO, Robert C.
;    APPLICANT:  WONG-STAAL, Flossie
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF HTLV-III DNA
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morgan & Finnegan, L.L.P.
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10154-0053
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,210
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/693,866
;      FILING DATE:  23-JAN-1985
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/659,339
;      FILING DATE:  10-OCT-1984
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Serunian, Leslie A.
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4193US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1015 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HTLV-III
;    FEATURE:
;      NAME/KEY:  Protein
;      LOCATION:  1..1015
;      OTHER INFORMATION:  /note= "pol protein of HTLV-III"
US-08-463-210-9

Alignment Scores:
Pred. No.:              2.34e-206     Length:        1015
Score:                  3432.00       Matches:       650
Percent Similarity:     93.18%        Conservative:  33
Best Local Similarity:  88.68%        Mismatches:    30
Query Match:            75.35%        Indels:        20
DB:                     3             Gaps:          5

US-09-610-313B-32 (1-2457) x US-08-463-210-9 (1-1015)
```

```
Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy    268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db     21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy    304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db     41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy    358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db     61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy    412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db     81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy    472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db    101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy    532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db    121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy    592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db    141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy    652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db    161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy    712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db    181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy    772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db    201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy    832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db    221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy    892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db    241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy    952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db    261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy   1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db    281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy   1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db    301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy   1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db    321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy   1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db    341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy   1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db    361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy   1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAG 1359
Db    381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy   1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
Db    401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy   1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
Db    421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy   1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
Db    441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy   1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
Db    461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy   1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
Db    481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy   1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
Db    501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy   1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
Db    521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy   1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
Db    541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy   1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
Db    561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy   1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
Db    581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy   1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
Db    601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy   2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
Db    621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy   2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
Db    641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy   2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
Db    661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy   2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
Db    681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy   2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
Db    701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy   2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
```

```
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 4
US-09-124-900-3
; Sequence 3, Application US/09124900
; Patent No. 6268484
; GENERAL INFORMATION:
; APPLICANT: KATINGER, Hermann
; APPLICANT: BUCHACHER, Andrea
; APPLICANT: ERNST, Wolfgang
; APPLICANT: BALLAUN, Claudia
; APPLICANT: PURTSCHER, Martin
; APPLICANT: TRKOLA, Alexandra
; APPLICANT: PREDL, Renate
; APPLICANT: SCHMATZ, Christine
; APPLICANT: KLIMA, Annelies
; APPLICANT: STEINDL, Franz
; APPLICANT: MUSTER, Thomas
; TITLE OF INVENTION: HIV-Vaccines
; FILE REFERENCE: 1939-112P
; CURRENT APPLICATION NUMBER: US/09/124,900
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: PCT/EP95/01481
; PRIOR FILING DATE: 1995-04-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1015
; TYPE: PRT
; ORGANISM: Human imunodeficiency virus type 1
US-09-124-900-3

Alignment Scores:

Pred. No.:	2.34e-206	Length:	1015
Score:	3432.00	Matches:	650
Percent Similarity:	93.18%	Conservative:	33
Best Local Similarity:	88.68%	Mismatches:	30
Query Match:	75.35%	Indels:	20
DB:	3	Gaps:	5

US-09-610-313B-32 (1-2457) x US-09-124-900-3 (1-1015)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAG 1359
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
```

```
Qy       1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
              ||||||:::||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
              ||||||         ||||||||||||||||||||||||||||||||||:::|||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
              |||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
              ||||||||||||||||::::::||||||||||||:::|||   ||||||   ||||||
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
              ||||||||||||||||||||||||||    ||||||||||||||||||||   ||||||
Db        641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
              :::||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
                 ||||||||||:::|||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 5
US-08-463-028-9
; Sequence 9, Application US/08463028
; Patent No. 6610476
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAAL, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,028
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1015
; OTHER INFORMATION: /note= "pol protein of HTLV-III"
US-08-463-028-9

Alignment Scores:

Pred. No.:	2.34e-206	Length:	1015
Score:	3432.00	Matches:	650
Percent Similarity:	93.18%	Conservative:	33
Best Local Similarity:	88.68%	Mismatches:	30
Query Match:	75.35%	Indels:	20
DB:	4	Gaps:	5

US-09-610-313B-32 (1-2457) x US-08-463-028-9 (1-1015)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
              ||||||||||||||||||||||||    ||||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
                                      ||||||||||   ||||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||   :::|||    ||||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      ||||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              |||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::||||||||||||||||||||||||||||||||||||:::|||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
```

```
Qy    712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db    181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy    772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db    201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy    832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db    221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy    892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db    241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy    952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db    261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy   1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db    281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy   1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db    301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy   1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db    321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy   1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db    341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy   1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db    361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy   1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAG 1359
Db    381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy   1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
Db    401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy   1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
Db    421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy   1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
Db    441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy   1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
Db    461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy   1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
Db    481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy   1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
Db    501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy   1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
Db    521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy   1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
Db    541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy   1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
Db    561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy   1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
Db    581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy   1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
Db    601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy   2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
Db    621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy   2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
Db    641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy   2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
Db    661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy   2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
Db    681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy   2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
Db    701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy   2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db    721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 6
US-07-743-357-2
; Sequence 2, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900

```
;      TELEFAX:  (613) 237-0045
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1016 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  protein
;     HYPOTHETICAL:  NO
;     FRAGMENT TYPE:  internal
;     ORIGINAL SOURCE:
;       ORGANISM:  Human immunodeficiency virus type 1
;       STRAIN:  BH102
US-07-743-357-2

Alignment Scores:
Pred. No.:              2.34e-206     Length:         1016
Score:                  3432.00       Matches:        650
Percent Similarity:     93.18%        Conservative:   33
Best Local Similarity:  88.68%        Mismatches:     30
Query Match:            75.35%        Indels:         20
DB:                     2             Gaps:           5

US-09-610-313B-32 (1-2457) x US-07-743-357-2 (1-1016)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy    268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db     21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy    304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db     41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy    358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db     61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy    412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db     81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy    472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db    101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy    532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db    121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy    592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db    141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy    652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db    161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy    712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db    181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy    772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db    201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy    832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db    221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy    892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db    241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy    952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db    261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy   1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db    281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy   1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db    301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy   1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db    321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy   1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db    341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy   1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db    361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy   1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAG 1359
Db    381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy   1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
Db    401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy   1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
Db    421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy   1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
Db    441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy   1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
Db    461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy   1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
Db    481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy   1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
Db    501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy   1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
Db    521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy   1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
Db    541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy   1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
Db    561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy   1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
Db    581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy   1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
```

```
                 |||:::|||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db         601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy        2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
               ||||||||||||||||:::::::||||||||||||:::|||    ||||||    ||||||
Db         621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy        2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
               ||||||||||||||||||||||||||||    ||||||||||||||||||    |||||||
Db         641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy        2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy        2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy        2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
               :::|||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db         701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy        2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
                  ||||||||||:::||||||||||||||||||
Db         721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 7
US-09-309-572-17
; Sequence 17, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
;   FEATURE:
;   OTHER INFORMATION: pol polyprotein
US-09-309-572-17

```
Alignment Scores:
Pred. No.:              3.6e-206      Length:         1003
Score:                  3429.00       Matches:        648
Percent Similarity:     94.59%        Conservative:   34
Best Local Similarity:  89.88%        Mismatches:     31
Query Match:            75.28%        Indels:         8
DB:                     4             Gaps:           4

US-09-610-313B-32 (1-2457) x US-09-309-572-17 (1-1003)

Qy         220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
               ||||||||||||||||||||||||||||||||||||||||||||||||    ||||||||
Db           1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy         280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                  ||||||||||||||||||    |||||||||||||||      |||    :::|||
Db          21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy         334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                  ||||||||||||||:::||||||||||||:::      :::|||||||||||||||
Db          41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy         388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
               ||||||||||||||||||||:::||||||:::|||||||||:::|||||||||||||||
Db          61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy         448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
               |||||||||||||||||||||||||||||||||:::|||||||||:::|||||||||
Db          81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy         508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
               ||||||||||||||||||||||||||||||||||||    ||||||||||||||||||||
Db         101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120

Qy         568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
               ||||||||||||    |||||||||||||||||||:::|||||||||||||||||||||
Db         121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy         628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
               ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db         141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy         688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy         748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
               ||||||||||||||||||||||||||||      ||||||    ||||||||||||||||||
Db         181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy         808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
               ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db         201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy         868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy         928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
               |||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db         241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSer 260

Qy         988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
               |||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db         261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy        1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy        1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
               ||||||||||||||||||||||||||||||||||||:::|||||||||    ||||||||
Db         301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320

Qy        1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
               ||||||||||||||||||||    :::||||||:::|||||||||||||||
Db         321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy        1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                  ||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db         341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy        1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
               |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy        1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
               |||||||||         ||||||||||||||||||||||||||||||||||||||||
Db         381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy        1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
               ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db         401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy        1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
               ||||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db         421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
```

```
Qy       1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db        441 ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520

Qy       1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy       1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560

Qy       1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db        581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy       2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db        621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
Db        701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2356 GAT 2358
Db        721 Asp 721
```

RESULT 8
US-09-718-096-17
; Sequence 17, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718,096
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: DE 19856463.5
; PRIOR FILING DATE: 1998-11-26
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
;   FEATURE:
;   OTHER INFORMATION: pol polyprotein
US-09-718-096-17

```
Alignment Scores:
Pred. No.:              3.6e-206       Length:         1003
Score:                  3429.00        Matches:        648
Percent Similarity:     94.59%         Conservative:   34
Best Local Similarity:  89.88%         Mismatches:     31
Query Match:            75.28%         Indels:         8
DB:                     4              Gaps:           4

US-09-610-313B-32 (1-2457) x US-09-718-096-17 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
```

```
                |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||||||:::|||||||||   ||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              ||||||||||||||||||    :::||||||:::||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
              |||||||||      ||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
              ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
                 |||||||||||::::::||||||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
              ||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
              :::|||||||||||||||||||||         |||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
              ||||||||||||||||||||||||||||||:::|||:::   ||||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520

Qy       1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
              |||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy       1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
              ||||||||||:::|||||||||||||||||||||||||||   |||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560

Qy       1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
              ||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy       2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
              :::|||   ||||||:::||||||||||||||||||||||||||||||   ||||||
Db        621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
              |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
              ||||||:::|||   |||||||||   ||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2356 GAT 2358
              |||
Db        721 Asp 721
```

RESULT 9
US-07-743-357-5
; Sequence 5, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
;   APPLICANT: Kang, Yong C.
;   TITLE OF INVENTION: Polypeptide having immunological
;   TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: KIRBY EADES GALE BAKER
;   STREET: Box 3432, Station D
;   CITY: Ottawa
;   STATE: Ontario
;   COUNTRY: Canada
;   ZIP: K1M 1H8
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/743,357
;   FILING DATE: 21-AUG-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/CA90/00062
;   FILING DATE: 23-FEB-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Gale, Edwin J.
;   REGISTRATION NUMBER: 28,584
;   REFERENCE/DOCKET NUMBER: 30924-2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (613) 237-6900
;   TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1016 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   FRAGMENT TYPE: internal
;   ORIGINAL SOURCE:
;   ORGANISM: Human immunodeficiency virus type 1
;   STRAIN: BRU

```
US-07-743-357-5

Alignment Scores:
Pred. No.:              5.56e-206     Length:         1016
Score:                  3426.00       Matches:        648
Percent Similarity:     93.32%        Conservative:   36
Best Local Similarity:  88.40%        Mismatches:     29
Query Match:            75.21%        Indels:         20
DB:                     2             Gaps:           5

US-09-610-313B-32 (1-2457) x US-07-743-357-5 (1-1016)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy    268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db     21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy    304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db     41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArg 60

Qy    358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db     61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy    412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db     81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy    472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db    101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy    532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db    121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy    592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db    141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy    652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db    161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy    712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db    181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy    772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db    201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy    832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db    221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy    892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db    241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy    952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db    261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy   1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db    281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy   1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db    301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy   1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db    321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340

Qy   1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db    341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy   1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db    361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy   1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAG 1359
Db    381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy   1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
Db    401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy   1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
Db    421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy   1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
Db    441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy   1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
Db    461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy   1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
Db    481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy   1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
Db    501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy   1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
Db    521 TyrAlaArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy   1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
Db    541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy   1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
Db    561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy   1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
Db    581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy   1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
Db    601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaSerArgGluThrLysLeuGly 620

Qy   2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
Db    621 LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValValThrLeuThrAspThrThr 640

Qy   2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
Db    641 AsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy   2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
```

```
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
              :::|||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
                 ||||||||||||:::||||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 10
US-07-743-357-3
; Sequence 3, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1016 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  BH5
US-07-743-357-3

Alignment Scores:
Pred. No.:              6.41e-206      Length:         1016
Score:                  3425.00        Matches:        648
Percent Similarity:     93.18%         Conservative:   35
Best Local Similarity:  88.40%         Mismatches:     30
Query Match:            75.19%         Indels:         20
DB:                     2              Gaps:           5

US-09-610-313B-32 (1-2457) x US-07-743-357-3 (1-1016)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
              ||||||||||||||||||||||||    ||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
                                      |||||||||   ||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||    :::|||    |||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      ||||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              ||||||||||||||||||||||||||||||||||||||||||||||||    ||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::||||||||||||||||||||||||||||||||||:::||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
              |||       ||||||   ||||||||||||||||||||||:::|||||||||||||||
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        241 LeuValAspPheArgGluLeuAsnArgArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
              |||||||||||||||||||||||         ||||||||||||||||||||||||||||
Db        301 SerIleAsnAsnGluThrProGlySerGlyTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
              :::||||||:::|||||||||||||||            ||||||||||||||||||||
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
```

```
                 |||||||||||||||||  |||||||||||||||||:::||||||||||||||||||
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1306 TTCACCACCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCCTG------CCCATCGAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        381 PheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
              ||||||||||||||||||||:::|||||||||  ||||||||||||:::||||||||
Db        401 LeuHisProAspLysTrpThrIleGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
              ||||||||||||||||||||||||||||||||||||   |||||||||||:::::::::
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
              |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
              ||||||:::||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
              ||||||      ||||||||||||||||||||||||||||||||||:::|||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
              |||:::|||||||||||||||||||||||||||||||:::|||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaSerArgGluThrLysLeuGly 620

Qy       2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
              ||||||||||||||||||:::|||||||||||||||:::|||:::||||||   ||||||
Db        621 LysAlaGlyTyrValThrAsnArgGlyArgGlnLysValValThrLeuThrHisThrThr 640

Qy       2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
              |||||||||||||||||||||||||||||||||   ||||||||||||||||||  ||||||
Db        641 AsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
              :::|||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
                 ||||||||||||:::||||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 11
US-07-743-357-4
; Sequence 4, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1016 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  PV22
US-07-743-357-4

Alignment Scores:
Pred. No.:                9.87e-206      Length:        1016
Score:                    3422.00        Matches:       647
Percent Similarity:       93.18%         Conservative:  36
Best Local Similarity:    88.27%         Mismatches:    30
Query Match:              75.13%         Indels:        20
DB:                       2              Gaps:          5

US-09-610-313B-32 (1-2457) x US-07-743-357-4 (1-1016)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
              ||||||||||||||||||||||||   |||||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
                                      |||||||||   ||||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||   :::|||   ||||||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
```

```
Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      |||||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::|||||||||||||||||||||||||||||||||:::|||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
              |||      ||||||   ||||||||||||||||||||||||:::|||||||||||||||
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||   
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
              :::||||||:::|||||||||||||||            |||||||||||||||||||||
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAG 1359
                 ||||||||||||||||||||||||||||||||||||||||||            |||
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
              |||||||||||||||||||||||||||||||||   |||||||||||||||:::||||||
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
              ||||||||||||||||||||||||||||||||||||   ||||||||||||:::::::::
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
              ||||||:::||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
              ||||||      ||||||||||||||||||||||||||||||||||||:::|||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
              |||:::|||||||||||||||||||||||||||||||||||||||||||||::::::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGly 620

Qy       2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
              ||||||||||||:::|||::::::||||||||||||:::|||   ||||||   ||||||
Db        621 LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
              |||||||||||||||||||||||||||   |||||||||||||||||||||   ||||||
Db        641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSer 680

Qy       2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlnLysValTyrLeu 700

Qy       2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
              :::|||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
                 ||||||||||||:::|||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

```
RESULT 12
US-07-743-357-10
; Sequence 10, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1003 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  ELI
US-07-743-357-10

Alignment Scores:
Pred. No.:               2.17e-205      Length:         1003
Score:                   3416.50        Matches:        642
Percent Similarity:      94.72%         Conservative:   40
Best Local Similarity:   89.17%         Mismatches:     31
Query Match:             75.01%         Indels:         7
DB:                      2              Gaps:           4

US-09-610-313B-32 (1-2457) x US-07-743-357-10 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||:::||||||||||||||||||||||||         |||      :::|||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerProLysGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
                 ||||||||||||||||||||||||||||:::|||   |||   ||||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTG 390
              |||:::   ||||||||||||||||||||:::          ||||||||||||||||||||
Db         41 SerLysThrGlyAlaGluArgGlnGlyThrValSerPheAsnPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              |||||||||||||||||:::|||||:::|||||||||:::|||||||||||||||||||
Db         61 TrpGlnArgProLeuValAlaIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||:::||||||||||||||||||||||:::|||||||||||||||||||||||
Db        121 IleCysGlyGlnLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              |||||||:::||||||||:::||||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              ||||||||||||||||||||||||   ||||||   :::||||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrAspMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||::::::||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        201 IleSerArgIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleSerSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
              ||||||||||||||||||     :::||||||||||:::||||||||||||
Db        321 IleLeuGluProPheArgLysGlnAsnProGluMetValIleTyrGlnTyrMetAspAsp 340

Qy       1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
              |||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLysLeuArg 360

Qy       1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
              :::||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuArgTrpGlyPheThrArgProAspLysLysHisGlnLysGluProPro 380

Qy       1345 TTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1398
              ||||||            ||||||||||||||||||||||||||||||   |||:::||||||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400

Qy       1399 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1458
              ||||||||||||||||||||||||||||||||||   ||||||   :::|||||||||
Db        401 ProGluLysGluSerTrpThrValAsnAspIleGlnAsnLeuValGluArgLeuAsnTrp 420

Qy       1459 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr 440

Qy       1519 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1578
              ||||||||||||:::::::::||||||||||||||||||||||||||||||||||||||||
Db        441 LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460

Qy       1579 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1638
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1639 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1698
```

```
                 ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy       1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1758
              |||||||||||||||||||||||||:::||||||   ||||||||||||||||||||
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520

Qy       1759 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1818
              ||||||   ||||||||||:::|||:::   ||||||||||||||||||:::|||
Db        521 GlnLeuAlaGluAlaValGlnArgIleSerThrGluSerIleValIleTrpGlyArgThr 540

Qy       1819 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878
              ||||||||||||||||||||||||||||||||||||||||||||||   :::||||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpAlaGluTyrTrp 560

Qy       1879 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1938
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1939 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 1998
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       1999 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2058
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||:::
Db        601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2059 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118
              |||    ||||||:::||||||||||||||||||||||||||||||||   ||||||||
Db        621 ValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640

Qy       2119 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2178
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2179 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2238
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIle 680

Qy       2239 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2298
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2299 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||:::||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 13
US-07-743-357-8
; Sequence 8, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
;   APPLICANT: Kang, Yong C.
;   TITLE OF INVENTION: Polypeptide having immunological
;   TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: KIRBY EADES GALE BAKER
;   STREET: Box 3432, Station D
;   CITY: Ottawa
;   STATE: Ontario
;   COUNTRY: Canada
;   ZIP: K1M 1H8
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/743,357
;   FILING DATE: 21-AUG-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/CA90/00062
;   FILING DATE: 23-FEB-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Gale, Edwin J.
;   REGISTRATION NUMBER: 28,584
;   REFERENCE/DOCKET NUMBER: 30924-2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (613) 237-6900
;   TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1003 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   FRAGMENT TYPE: internal
;   ORIGINAL SOURCE:
;   ORGANISM: Human immunodeficiency virus type 1
;   STRAIN: RF
US-07-743-357-8

```
Alignment Scores:
Pred. No.:              3.34e-205     Length:         1003
Score:                  3413.50       Matches:        643
Percent Similarity:     94.72%        Conservative:   39
Best Local Similarity:  89.31%        Mismatches:     31
Query Match:            74.94%        Indels:         7
DB:                     2             Gaps:           4

US-09-610-313B-32 (1-2457) x US-07-743-357-8 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::|||||||||||||||||||||||||||||        |||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluLeuSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
                 |||||||||||||||||   |||||||||||||||   |||   ||||||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSerLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTG 390
              ||||||||||||    :::||||||||||||:::      :::||||||||||||||||||
Db         41 SerGluAlaGlyGluAspArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||||:::|||::::::|||:::||||||||||:::|||||||||||||||||
Db         61 TrpGlnArgProIleValThrValLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||   |||||||||||||||||||||:::||||||||||||||||||||||||
Db        121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              |||||||||:::|||||||||:::||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              ||||||||||||||||||||||||      ||||||   ||||||||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLys 200
```

```
Qy    811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
          |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy    871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy    931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy    991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
          |||||||||||||||||||||||||||||||||||||||||||||::::::|||||||||
Db    261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysGluPheArgLys 280

Qy   1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
          ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db    281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProArgIleArgTyrGlnTyr 300

Qy   1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
          |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db    301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy   1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
          ||||||||||||||:::   :::|||||||||||||||||||||
Db    321 IleLeuGluProPheLysLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy   1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
          ||||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db    341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgIleLysIleGluGluLeuArg 360

Qy   1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
          :::||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy   1345 TTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1398
          ||||||                                                   |||
Db    381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleValLeu 400

Qy   1399 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1458
          |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db    401 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy   1459 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1518
          |||||||||||||||   ||||||||||||:::||||||||||||||||||||||||
Db    421 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 440

Qy   1519 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1578
          ||||||||||||:::::|||   ||||||:::||||||||||||||||||||||||||||
Db    441 LysAlaLeuThrGluValValGlnLeuThrLysGluAlaGluLeuGluLeuAlaGluAsn 460

Qy   1579 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1638
          ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||:::
Db    461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy   1639 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1698
          |||||||||||||||||||||         ||||||||||||||||||||||||||||||
Db    481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy   1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1758
          |||||||||||||||||||||||||||:::||||||   |||||||||||||||||||||
Db    501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520

Qy   1759 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1818
          ·|||||||||||||||||||||||:::|||   |||||||||||||||||||||||||||
Db    521 GlnLeuThrGluAlaValGlnLysValAlaThrGluSerIleValIleTrpGlyLysThr 540

Qy   1819 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878
          ||||||||:::|||||||||||||||||||||||||||   |||||||||:::||||||
Db    541 ProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy   1879 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy   1939 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 1998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy   1999 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2058
          |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||:::
Db    601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy   2059 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118
          ||||||||||||:::||||||||||||||||||||||||||||||||   |||||||||
Db    621 ValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy   2119 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2178
          ||||||||||||      ||||||||||||||||||||||||||||||||||||||||||
Db    641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy   2179 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2238
          ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db    661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680

Qy   2239 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2298
          ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db    681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy   2299 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
          |||:::|||:::|||||||||   ||||||||||||||||||||||||||||||||||||
Db    701 GlnValAspArgLeuValSerThrGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 14
US-07-743-357-7

```
; Sequence 7, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1004 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
```

```
;     HYPOTHETICAL:  NO
;     FRAGMENT TYPE:  internal
;     ORIGINAL SOURCE:
;       ORGANISM:  Human immunodeficiency virus type 1
;       STRAIN:  SF2
US-07-743-357-7

Alignment Scores:
Pred. No.:              9.82e-205      Length:        1004
Score:                  3406.00        Matches:       642
Percent Similarity:     94.31%         Conservative:  38
Best Local Similarity:  89.04%         Mismatches:    33
Query Match:            74.77%         Indels:        8
DB:                     2              Gaps:          4

US-09-610-313B-32 (1-2457) x US-07-743-357-7 (1-1004)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy    280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------GACAACCCC 333
Db     21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40

Qy    334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db     41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy    388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db     61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy    448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db     81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100

Qy    508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db    101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProVal 120

Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db    321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy   1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db    341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy   1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db    361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy   1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
Db    381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMet 400

Qy   1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db    401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy   1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db    421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy   1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db    441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy   1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db    461 AsnArgGluIleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeu 480

Qy   1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db    481 ValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy   1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db    501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy   1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db    521 LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys 540

Qy   1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db    541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyr 560

Qy   1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db    561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy   1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db    581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy   1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db    601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy   2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db    621 ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy   2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
```

```
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720

Qy       2356 GAT 2358
Db        721 Asp 721

RESULT 15
US-09-319-588C-6
; Sequence 6, Application US/09319588C
; Patent No. 6509018
; GENERAL INFORMATION:
;  APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-INSERM
;  APPLICANT:  ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
;  APPLICANT:  INSTITUT PASTEUR
;  APPLICANT:  MAUCLERE, Philippe
;  APPLICANT:  LOUSSERT-AJAKA, Ibtissam
;  APPLICANT:  SIMON, Francois
;  APPLICANT:  SARAGOSTI, Sentob
;  APPLICANT:  BARRE-SINOUSSI, Francoise
;  TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
;  FILE REFERENCE: 598US12
;  CURRENT APPLICATION NUMBER: US/09/319,588C
;  CURRENT FILING DATE:  1999-08-27
;  PRIOR APPLICATION NUMBER: FR96/15087
;  PRIOR FILING DATE: 1996-12-09
;  NUMBER OF SEQ ID NOS: 98
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1014
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-09-319-588C-6

Alignment Scores:
Pred. No.:               3.72e-188    Length:          1014
Score:                   3140.50      Matches:         579
Percent Similarity:      89.70%       Conservative:    74
Best Local Similarity:   79.53%       Mismatches:      60
Query Match:             68.95%       Indels:          15
DB:                      4            Gaps:            5

US-09-610-313B-32 (1-2457) x US-09-319-588C-6 (1-1014)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProProAspAsn 20

Qy        280 AAC------CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG--------------- 318
Db         21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40

Qy        319 ---CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG---------GGCACC 366
Db         41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db         61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db         81 GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluGluLeuGln 100

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db        101 LeuGluGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArg 120

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db        121 GlnTyrAspAsnIleThrValAspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db        141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db        161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db        181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db        201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db        221 ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db        241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db        261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db        281 ProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGlu 300

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db        301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db        321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340

Qy       1207 GTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db        341 IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db        361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAG 1374
Db        381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
Db        401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
Db        421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
```

```
Db        441 LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||:::|||||||||:::||||||||||||
Db        461 AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              |||||||||   |||:::|||||||||||||||||||||||||||      |||||||||
Db        481 TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThr 500

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
              ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        501 TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
              :::|||||||||||||||:::|||||||||   |||   |||:::|||:::|||   |||
Db        521 SerAlaHisThrAsnAspIleLysGlnLeuValGluValValArgLysValAlaThrGlu 540

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||   |||
Db        541 SerIleValIleTrpGlyLysThrProLysPheArgLeuProValGlnLysGluValTrp 560

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
              |||   ||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        561 GluAlaTrpTrpThrAspHisTrpGlnAlaThrTrpIleProGluTrpGluPheValAsn 580

Qy       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
              ||||||||||||||||||||||||||||||||||||   |||||||||   |||||||||
Db        581 ThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGluProIleSerGlyAlaGlu 600

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
              |||||||||||||||||||||||||||||||||||||||:::|||||||||||||:::|||
Db        601 ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
              |||||||||||||||||||||:::||||||:::   :::|||||||||||||||   |||
Db        621 ThrAspArgGlyArgGlnLysValValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
              ||||||||||||   :::|||||||||:::||||||   :::||||||||||||||||||
Db        641 LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
              ||||||||||||:::|||||||||   :::||||||||||||||||||||||||||||||
Db        661 SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
              :::||||||||||||:::|||||||||||||||:::||||||||||||||||||||||||
Db        681 SerGlnIleIleGluGluLeuIleLysLysGluArgValTyrLeuSerTrpValProAla 700

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
              |||||||||||||||||||||||||||:::|||||||||||||||   ||||||||||||
Db        701 HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720

Qy       2335 GTGCTGTTCCTGGACGGCATCGAT 2358
              :::|||||||||||||||||||:::
Db        721 IleLeuPheLeuAspGlyIleGlu 728
```

Search completed: June 2, 2005, 05:13:50
Job time : 106.917 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 2, 2005, 03:13:23 ; Search time 34.0828 Seconds
(without alignments)
10815.332 Million cell updates/sec

Title: US-09-610-313B-30
Perfect score: 4583
Sequence: 1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2469

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
```
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155137_15173/app_query.fasta_1.7893
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_101_@runat_31052005_155137_15173 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3535	77.1	1005	2	US-07-743-357-1	Sequence 1, Appli
2	3531.5	77.1	1003	2	US-07-743-357-9	Sequence 9, Appli
3	3513	76.7	1015	3	US-08-463-210-9	Sequence 9, Appli
4	3513	76.7	1015	3	US-09-124-900-3	Sequence 3, Appli
5	3513	76.7	1015	4	US-08-463-028-9	Sequence 9, Appli
6	3513	76.7	1016	2	US-07-743-357-2	Sequence 2, Appli
7	3510	76.6	1003	4	US-09-309-572-17	Sequence 17, Appl
8	3510	76.6	1003	4	US-09-718-096-17	Sequence 17, Appl
9	3507	76.5	1016	2	US-07-743-357-5	Sequence 5, Appli
10	3506	76.5	1016	2	US-07-743-357-3	Sequence 3, Appli
11	3503	76.4	1016	2	US-07-743-357-4	Sequence 4, Appli
12	3497.5	76.3	1003	2	US-07-743-357-10	Sequence 10, Appl
13	3494.5	76.2	1003	2	US-07-743-357-8	Sequence 8, Appli
14	3487	76.1	1004	2	US-07-743-357-7	Sequence 7, Appli
15	3221.5	70.3	1014	4	US-09-319-588C-6	Sequence 6, Appli
16	3179	69.4	913	2	US-07-743-357-22	Sequence 22, Appl
17	3135.5	68.4	1350	4	US-09-952-060-35	Sequence 35, Appl
18	3116	68.0	913	2	US-07-743-357-6	Sequence 6, Appli
19	2860.5	62.4	875	4	US-09-952-060-6	Sequence 6, Appli
20	2855	62.3	850	4	US-09-952-060-2	Sequence 2, Appli
21	2827	61.7	562	3	US-09-117-217-14	Sequence 14, Appl
22	2827	61.7	562	4	US-09-735-487-14	Sequence 14, Appl
23	2814.5	61.4	875	4	US-09-952-060-8	Sequence 8, Appli
24	2809	61.3	850	4	US-09-952-060-4	Sequence 4, Appli
25	2796	61.0	560	4	US-10-205-641-1	Sequence 1, Appli
26	2545.5	55.5	1018	4	US-09-206-551-46	Sequence 46, Appl
27	2521.5	55.0	1016	4	US-09-625-972-24	Sequence 24, Appl
28	2270	49.5	441	4	US-09-991-258-16	Sequence 16, Appl
29	2247	49.0	1055	2	US-08-659-251-5	Sequence 5, Appli
30	2247	49.0	1055	3	US-09-256-490-5	Sequence 5, Appli
31	2247	49.0	1055	5	PCT-US96-11445-5	Sequence 5, Appli
32	2228.5	48.6	3080	6	5223423-4	Patent No. 5223423
33	2228.5	48.6	3080	6	5223423-4	Patent No. 5223423
34	2192	47.8	427	3	US-09-690-265-1	Sequence 1, Appli
35	1593	34.8	327	3	US-08-679-493A-71	Sequence 71, Appl
36	1590	34.7	314	1	US-08-589-446-6	Sequence 6, Appli
37	1590	34.7	314	1	US-08-444-882-6	Sequence 6, Appli
38	1590	34.7	314	2	US-08-389-459A-6	Sequence 6, Appli
39	1590	34.7	314	3	US-08-987-867A-6	Sequence 6, Appli
40	1576	34.4	327	3	US-08-679-493A-72	Sequence 72, Appl
41	1349	29.4	1150	3	US-09-238-303-9	Sequence 9, Appli
42	1349	29.4	1150	4	US-09-946-239-9	Sequence 9, Appli
43	1335	29.1	261	6	5320958-3	Patent No. 5320958
44	1335	29.1	261	6	5320958-3	Patent No. 5320958
45	1320	28.8	1031	3	US-08-811-682-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-07-743-357-1
; Sequence 1, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 1:

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1005 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  HXB2
US-07-743-357-1

Alignment Scores:
Pred. No.:               3.5e-213     Length:         1005
Score:                   3535.00      Matches:        659
Percent Similarity:      96.12%       Conservative:   34
Best Local Similarity:   91.40%       Mismatches:     24
Query Match:             77.13%       Indels:         4
DB:                      2            Gaps:           2

US-09-610-313B-30 (1-2469) x US-07-743-357-1 (1-1005)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||    |||||||||||||||||||   |||||||||
Db          2 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 21

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                 ||||||||||||||||||    ||||||||||||||      |||   :::|||
Db         22 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 41

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 |||||||||||||:::||||||||||||:::        ||||||||||:::|||
Db         42 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 61

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||:::||||||:::||||||||:::||||||||||||||
Db         62 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 81

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db         82 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 101

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 121

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              |||||||||||    ||||||||||||||||||||:::||||||||||||||||||||
Db        122 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 141

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              |||||||||||:::|||||||||:::||||||||||||||||||||||||||||||||
Db        142 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 161

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 181

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||||      ||||||   |||||||||||||||||
Db        182 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 201

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 221

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 241

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 261

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 281

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 301

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        302 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 321

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
              ||||||||||||||||||||    :::||||||:::||||||||||||||||||||||
Db        322 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 341

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
              |||||||||||||||||||||||||||||||||||||      |||||||||||||||
Db        342 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 361

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
              |||:::|||||||||||||||||    ||||||||||||||||||||||||||||||||
Db        362 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 381

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 401

Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
              |||||||||||:::||||||||||:::|||||||||||||||||||||||||||||||||
Db        402 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 421

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 441

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
                 ||||||||||:::::::::|||||||||||||||||||||||||||||||||||
Db        442 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 461

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
              |||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        462 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 481

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
              :::||||||||||||||||||||||      |||||||||||||||||||||||||||||
Db        482 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 501

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
              ||||||||||||||||||||||||||||:::||||||   |||||||||||||||||
Db        502 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 521

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
              ||||||||||||||||||||||||||||      ||||||||||||||||||||||||
Db        522 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 541

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
              ||||||||||:::||||||||||||||||||||||||||||||||||||||:::||||
Db        542 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 561

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        562 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 581

Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db        582 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 601

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
              ||||||||||||||||:::|||||||||||||||||||||:::|||||||||||||||
Db        602 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 621
```

```
Qy       2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
              :::|||:::|||||:::||||||||||||||||||||||||||||||   |  ||||||
Db        622 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 641

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        642 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 661

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        662 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 681

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        682 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 701

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
              ||||||:::|||||||||||||||   |||||||||||||||||||||||||||||||||
Db        702 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 721

Qy       2368 GAT 2370
              |||
Db        722 Asp 722

RESULT 2
US-07-743-357-9
; Sequence 9, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1003 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  MAL
US-07-743-357-9

  Alignment Scores:
  Pred. No.:               5.8e-213      Length:          1003
  Score:                   3531.50       Matches:         656
  Percent Similarity:      96.53%        Conservative:    39
  Best Local Similarity:   91.11%        Mismatches:      22
  Query Match:             77.06%        Indels:          3
  DB:                      2             Gaps:            2

US-09-610-313B-30 (1-2469) x US-07-743-357-9 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
                 ||||||||||||||||||||||||||||||:::|||         ||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACCCTG 390
              ||||||   ||||||||||||||||||         :::   :::||||||||||||||||
Db         41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPheSerPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||||:::|||:::::::::||||||||||||:::||||||||||||||||||
Db         61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              ||||||||||||||||||||||||||||||::::::||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||||||||||||||||||:::||||:::|||||||||||||||||||||||||
Db        121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              ||||||||||||||||||||||||||     ||||||::::::|||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
```

```
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230
              ||||||||||||||||    :::|||||||||||||||||||||||||||||||||||||
Db        321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy       1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy       1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350
              :::|||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy       1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1410
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400

Qy       1411 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1470
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy       1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440

Qy       1531 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1590
              |||||||||||||||||||||||||||||    |||||||||||||||||||||||||||
Db        441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460

Qy       1591 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1650
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1651 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1710
              |||||||||||||||||||||      |||||||||||||||||||||||||||    :::
Db        481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500

Qy       1711 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1770
              |||||||||||||||||||||||||||:::::::::::|||||||||||||||||||||||
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520

Qy       1771 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1830
              |||||||||||||||||||||||||||||||    |||||||||||||||||||||||||
Db        521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540

Qy       1831 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890
              ||||||||||||||||||||||||||||||||||||||||    ||||||||:::||||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy       1891 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1951 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2010
              ||||||||||||    |||||||||:::|||||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       2011 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2070
              |||||||||||||||    |||||||||||||||||||||||||||||||||||||:::
Db        601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2071 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
              |||||||||||||||||||||||||||||||||||||||||||||||||    |||||||||
Db        621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy       2131 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2190
              ||||||||||||    ||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2191 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2250
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

Qy       2251 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2310
              :::|||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              |||:::||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 3
US-08-463-210-9

```
; Sequence 9, Application US/08463210
; Patent No. 6001977
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, Nancy T.
;    APPLICANT:  GALLO, Robert C.
;    APPLICANT:  WONG-STAAL, Flossie
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF HTLV-III DNA
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morgan & Finnegan, L.L.P.
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10154-0053
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,210
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/693,866
;      FILING DATE:  23-JAN-1985
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/659,339
;      FILING DATE:  10-OCT-1984
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Serunian, Leslie A.
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4193US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1015 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HTLV-III
;    FEATURE:
;      NAME/KEY:  Protein
;      LOCATION:  1..1015
;      OTHER INFORMATION:  /note= "pol protein of HTLV-III"
US-08-463-210-9
```

Alignment Scores:

Pred. No.:	8.35e-212	Length:	1015
Score:	3513.00	Matches:	658
Percent Similarity:	94.27%	Conservative:	33
Best Local Similarity:	89.77%	Mismatches:	26
Query Match:	76.65%	Indels:	16
DB:	3	Gaps:	3

US-09-610-313B-30 (1-2469) x US-08-463-210-9 (1-1015)

```
Qy      220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC----------- 267
Db        1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy      268 -----------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db       21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy      304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db       41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy      358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db       61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy      412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db       81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy      472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db      101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy      532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db      121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy      592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db      141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy      652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db      161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy      712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db      181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy      772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db      201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy      832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db      221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy      892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db      241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy      952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db      261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy     1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db      281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy     1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db      301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy     1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db      321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy     1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
Db      341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy     1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
Db      361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy     1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
Db      381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy     1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
Db      401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy     1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
Db      421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy     1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
Db      441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy     1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
Db      461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy     1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
Db      481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy     1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
Db      501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy     1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
Db      521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy     1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
Db      541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy     1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
Db      561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy     1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
Db      581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy     1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
Db      601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy     2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
Db      621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy     2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
Db      641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy     2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
Db      661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy     2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
Db      681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy     2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
Db      701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy     2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
```

```
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 4
US-09-124-900-3
; Sequence 3, Application US/09124900
; Patent No. 6268484
; GENERAL INFORMATION:
;  APPLICANT:  KATINGER, Hermann
;  APPLICANT:  BUCHACHER, Andrea
;  APPLICANT:  ERNST, Wolfgang
;  APPLICANT:  BALLAUN, Claudia
;  APPLICANT:  PURTSCHER, Martin
;  APPLICANT:  TRKOLA, Alexandra
;  APPLICANT:  PREDL, Renate
;  APPLICANT:  SCHMATZ, Christine
;  APPLICANT:  KLIMA, Annelies
;  APPLICANT:  STEINDL, Franz
;  APPLICANT:  MUSTER, Thomas
;  TITLE OF INVENTION:   HIV-Vaccines
;  FILE REFERENCE:   1939-112P
;  CURRENT APPLICATION NUMBER: US/09/124,900
;  CURRENT FILING DATE:  1998-07-30
;  PRIOR APPLICATION NUMBER:    PCT/EP95/01481
;  PRIOR FILING DATE:    1995-04-19
;  NUMBER OF SEQ ID NOS:   11
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 1015
;   TYPE: PRT
;   ORGANISM: Human imunodeficiency virus type 1
US-09-124-900-3

Alignment Scores:
Pred. No.:               8.35e-212       Length:        1015
Score:                   3513.00         Matches:       658
Percent Similarity:      94.27%          Conservative:  33
Best Local Similarity:   89.77%          Mismatches:    26
Query Match:             76.65%          Indels:        16
DB:                      3               Gaps:          3

US-09-610-313B-30 (1-2469) x US-09-124-900-3 (1-1015)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
```

```
Qy       1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
              ||||||:::||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
              ||||||      ||||||||||||||||||||||||||||||||||:::|||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
              |||:::||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
              ||||||||||||||||||::::::||||||||||||:::|||   ||||||   ||||||
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
              |||||||||||||||||||||||||||   |||||||||||||||||||||   ||||||
Db        641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
              :::|||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
                 ||||||||||||:::||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

```
RESULT 5
US-08-463-028-9
; Sequence 9, Application US/08463028
; Patent No. 6610476
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, Nancy T.
;    APPLICANT:  GALLO, Robert C.
;    APPLICANT:  WONG-STAAL, Flossie
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF HTLV-III DNA
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morgan & Finnegan, L.L.P.
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10154-0053
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,028
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/693,866
;      FILING DATE:  23-JAN-1985
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/659,339
;      FILING DATE:  10-OCT-1984
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Serunian, Leslie A.
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4193US3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1015 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HTLV-III
;    FEATURE:
;      NAME/KEY:  Protein
;      LOCATION:  1..1015
;      OTHER INFORMATION:  /note= "pol protein of HTLV-III"
US-08-463-028-9

Alignment Scores:
Pred. No.:              8.35e-212      Length:        1015
Score:                  3513.00        Matches:       658
Percent Similarity:     94.27%         Conservative:  33
Best Local Similarity:  89.77%         Mismatches:    26
Query Match:            76.65%         Indels:        16
DB:                     4              Gaps:          3

US-09-610-313B-30 (1-2469) x US-08-463-028-9 (1-1015)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
              ||||||||||||||||||||||||    ||||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
                                      |||||||||    ||||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||   :::|||   ||||||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      ||||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::||||||||||||||||||||||||||||||||||:::|||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
```

```
Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
              |||      ||||||   ||||||||||||||||||||:::|||||||||||||||||||
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
              ||||||||||:::||||||||||||||||||||||||||||||||||||||:::
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
              :::||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
              ||||||||||||||||||   ||||||||||||||||:::||||||||||||||||||||
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
              |||||||||||||||||||||||||||||||   |||||||||||:::||||||||||||
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
              ||||||||||||||||||||||||||||||||||||   ||||||||||||:::::::::
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
                       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
              ||||||:::||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
              ||||||         |||||||||||||||||||||||||||||||||||:::||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
              ||||||||||||||||:::::::||||||||||||:::|||   ||||||   ||||||
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
              |||||||||||||||||||||||||   |||||||||||||||||||||   ||||||
Db        641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
              :::|||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
                 ||||||||||||:::||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 6
US-07-743-357-2
; Sequence 2, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
```

```
;     TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1016 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  BH102
US-07-743-357-2

Alignment Scores:
Pred. No.:              8.36e-212    Length:        1016
Score:                  3513.00      Matches:       658
Percent Similarity:     94.27%       Conservative:  33
Best Local Similarity:  89.77%       Mismatches:    26
Query Match:            76.65%       Indels:        16
DB:                     2            Gaps:          3

US-09-610-313B-30 (1-2469) x US-07-743-357-2 (1-1016)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
```

```
                 |||:::||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db         601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy        2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
               ||||||||||||||||||:::::||||||||||||:::|||   ||||||  ||||||
Db         621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy        2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
               ||||||||||||||||||||||||||   |||||||||||||||||||   ||||||
Db         641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy        2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy        2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy        2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
               :::|||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db         701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy        2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
                  ||||||||||||:::||||||||||||||||||
Db         721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 7
US-09-309-572-17
; Sequence 17, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-309-572-17

Alignment Scores:

Pred. No.:	1.28e-211	Length:	1003
Score:	3510.00	Matches:	656
Percent Similarity:	95.70%	Conservative:	34
Best Local Similarity:	90.98%	Mismatches:	27
Query Match:	76.59%	Indels:	4
DB:	4	Gaps:	2

US-09-610-313B-30 (1-2469) x US-09-309-572-17 (1-1003)

```
Qy         220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
               ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db           1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy         280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                  ||||||||||||||||   |||||||||||||||      |||   :::|||
Db          21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy         334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                  ||||||||||||||:::||||||||||||:::      :::||||||||||||||
Db          41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy         388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
               |||||||||||||||||||||:::||||||:::|||||||||:::|||||||||||||||
Db          61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy         448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
               |||||||||||||||||||||||||||||||||:::|||||||||:::||||||||||
Db          81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy         508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
               ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db         101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120

Qy         568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
               |||||||||||   ||||||||||||||||||||:::|||||||||||||||||||||||
Db         121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy         628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
               |||||||||:::||||||||:::|||||||||||||||||||||||||||||||||||||
Db         141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy         688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy         748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
               ||||||||||||||||||||||||||      ||||||   ||||||||||||||||||
Db         181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy         808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
               ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db         201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy         868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy         928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
               |||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db         241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSer 260

Qy         988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
               ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db         261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy        1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy        1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
               ||||||||||||||||||||||||||||||||||||||:::|||||||||   ||||||||
Db         301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320

Qy        1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
               |||||||||||||||||||     :::||||||:::|||||||||||||||||||||||
Db         321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy        1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
               ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db         341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy        1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
               |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy        1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy        1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
               ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db         401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy        1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
               |||||||||||||||||    ||||||||||||||||||||||||||||||||||||||||
Db         421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
```

```
Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
              ||||||||||||::::::||||||||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
              :::|||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
              |||||||||||||||||||||||||||||||:::|||:::   |||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
              ||||||||||||:::||||||||||||||||||||||||||   |||||||||:::||||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy       2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
              :::|||   ||||||:::|||||||||||||||||||||||||||||||   ||||||||
Db        621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
              ||||||:::|||   |||||||||   |||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370
              |||
Db        721 Asp 721
```

RESULT 8
US-09-718-096-17
; Sequence 17, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718,096
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: DE 19856463.5
; PRIOR FILING DATE: 1998-11-26
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
;   FEATURE:
;   OTHER INFORMATION: pol polyprotein
US-09-718-096-17

```
Alignment Scores:
Pred. No.:               1.28e-211     Length:         1003
Score:                   3510.00       Matches:        656
Percent Similarity:      95.70%        Conservative:   34
Best Local Similarity:   90.98%        Mismatches:     27
Query Match:             76.59%        Indels:         4
DB:                      4             Gaps:           2
```

US-09-610-313B-30 (1-2469) x US-09-718-096-17 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
              ||||||||||||||||||   |||||||||||||||            |||   :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
              |||||||||||||||:::||||||||||||:::            :::||||||||||||
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||||:::||||||:::|||||||||:::|||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||||:::|||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              ||||||||||||   ||||||||||||||||||:::||||||||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||||      ||||||   |||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              ||||||||||||||||||||||||||||||||||||:::||||||||   ||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
              |||||||||||||||||||    :::||||||:::|||||||||||||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
              ||||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
                 |||||||||||:::::||||||||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
              :::|||||||||||||||||||||         |||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
              ||||||||||||||||||||||||||||||||:::|||:::     ||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
              ||||||||||||||||||||||||||||||||||    ||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
              |||||||||||:::||||||||||||||||||||||||||||||    ||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
              |||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy       2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
              :::|||   ||||||:::|||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
              |||||||||||||||    |||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
              ||||||:::|||   |||||||||   |||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370
              |||
Db        721 Asp 721
```

RESULT 9
US-07-743-357-5
; Sequence 5, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
;   APPLICANT: Kang, Yong C.
;   TITLE OF INVENTION: Polypeptide having immunological
;   TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: KIRBY EADES GALE BAKER
;   STREET: Box 3432, Station D
;   CITY: Ottawa
;   STATE: Ontario
;   COUNTRY: Canada
;   ZIP: K1M 1H8
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/743,357
;   FILING DATE: 21-AUG-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/CA90/00062
;   FILING DATE: 23-FEB-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Gale, Edwin J.
;   REGISTRATION NUMBER: 28,584
;   REFERENCE/DOCKET NUMBER: 30924-2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (613) 237-6900
;   TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1016 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   FRAGMENT TYPE: internal
;   ORIGINAL SOURCE:
;   ORGANISM: Human immunodeficiency virus type 1
;   STRAIN: BRU

```
US-07-743-357-5

Alignment Scores:
Pred. No.:                1.98e-211      Length:        1016
Score:                    3507.00        Matches:       656
Percent Similarity:       94.41%         Conservative:  36
Best Local Similarity:    89.50%         Mismatches:    25
Query Match:              76.52%         Indels:        16
DB:                       2              Gaps:          3

US-09-610-313B-30 (1-2469) x US-07-743-357-5 (1-1016)

Qy     220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db       1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy     268 -----------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db      21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy     304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db      41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArg 60

Qy     358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db      61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy     412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db      81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy     472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db     101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy     532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db     121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy     592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db     141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy     652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db     161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy     712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db     181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy     772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db     201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy     832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db     221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy     892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db     241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy     952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db     261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy    1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db     281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy    1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db     301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy    1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db     321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340

Qy    1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
Db     341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy    1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
Db     361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy    1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
Db     381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy    1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
Db     401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy    1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
Db     421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy    1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
Db     441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy    1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
Db     461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy    1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
Db     481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy    1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
Db     501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy    1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
Db     521 TyrAlaArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy    1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
Db     541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy    1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
Db     561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy    1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
Db     581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy    1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
Db     601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaSerArgGluThrLysLeuGly 620

Qy    2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
Db     621 LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValValThrLeuThrAspThrThr 640

Qy    2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
Db     641 AsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy    2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
```

```
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
              :::|||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
                 ||||||||||||:::|||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 10
US-07-743-357-3
; Sequence 3, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
;   APPLICANT: Kang, Yong C.
;   TITLE OF INVENTION: Polypeptide having immunological
;   TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: KIRBY EADES GALE BAKER
;   STREET: Box 3432, Station D
;   CITY: Ottawa
;   STATE: Ontario
;   COUNTRY: Canada
;   ZIP: K1M 1H8
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/743,357
;   FILING DATE: 21-AUG-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/CA90/00062
;   FILING DATE: 23-FEB-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Gale, Edwin J.
;   REGISTRATION NUMBER: 28,584
;   REFERENCE/DOCKET NUMBER: 30924-2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (613) 237-6900
;   TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1016 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   FRAGMENT TYPE: internal
;   ORIGINAL SOURCE:
;   ORGANISM: Human immunodeficiency virus type 1
;   STRAIN: BH5
US-07-743-357-3

Alignment Scores:			
Pred. No.:	2.29e-211	Length:	1016
Score:	3506.00	Matches:	656
Percent Similarity:	94.27%	Conservative:	35
Best Local Similarity:	89.50%	Mismatches:	26
Query Match:	76.50%	Indels:	16
DB:	2	Gaps:	3

US-09-610-313B-30 (1-2469) x US-07-743-357-3 (1-1016)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
              ||||||||||||||||||||||||   ||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 -----------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
                                     |||||||||   ||||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              ||||||||||||||         |||   :::|||      ||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      ||||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::|||||||||||||||||||||||||||||||||:::|||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
              |||      ||||||   |||||||||||||||||||||||:::|||||||||||||||
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        241 LeuValAspPheArgGluLeuAsnArgArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
              ||||||||||||||||||||||||         |||||||||||||||||||||||||||
Db        301 SerIleAsnAsnGluThrProGlySerGlyTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
              :::||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
```

```
                |||||||||||||||||   ||||||||||||||||:::||||||||||||||||
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 PheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
              ||||||||||||||||||:::|||||||||   ||||||||||:::|||||||||
Db        401 LeuHisProAspLysTrpThrIleGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
              ||||||||||||||||||||||||||||||||||||   |||||||||||::::::::
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
              |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
                       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
              |||||:::|||||   ||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
              |||||      ||||||||||||||||||||||||||||||||||||:::||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
              |||:::||||||||||||||||||||||||||||||:::||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaSerArgGluThrLysLeuGly 620

Qy       2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
              ||||||||||||||||:::||||||||||||||:::|||:::||||||   ||||||
Db        621 LysAlaGlyTyrValThrAsnArgGlyArgGlnLysValValThrLeuThrHisThrThr 640

Qy       2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
              ||||||||||||||||||||||||||   |||||||||||||||||||||   ||||||
Db        641 AsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
              :::|||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
                 |||||||||||:::||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 11
US-07-743-357-4
; Sequence 4, Application US/07743357
; Patent No. 5858646
;  GENERAL INFORMATION:
;    APPLICANT:  Kang, Yong C.
;    TITLE OF INVENTION:  Polypeptide having immunological
;    TITLE OF INVENTION:  activity for use as diagnostic reagent and/or vaccine
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KIRBY EADES GALE BAKER
;      STREET:  Box 3432, Station D
;      CITY:  Ottawa
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  K1M 1H8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/743,357
;      FILING DATE:  21-AUG-1991
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA90/00062
;      FILING DATE:  23-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gale, Edwin J.
;      REGISTRATION NUMBER:  28,584
;      REFERENCE/DOCKET NUMBER:  30924-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (613) 237-6900
;      TELEFAX:  (613) 237-0045
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1016 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  PV22
US-07-743-357-4

Alignment Scores:
Pred. No.:              3.53e-211      Length:         1016
Score:                  3503.00        Matches:        655
Percent Similarity:     94.27%         Conservative:   36
Best Local Similarity:  89.36%         Mismatches:     26
Query Match:            76.43%         Indels:         16
DB:                     2              Gaps:           3

US-09-610-313B-30 (1-2469) x US-07-743-357-4 (1-1016)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
              ||||||||||||||||||||||||   |||||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
                                      |||||||||   ||||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||   :::|||   |||||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
```

```
Qy      358 CAGGGCACCCTG------AACTTCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db       61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy      412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db       81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy      472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db      101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy      532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db      121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy      592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db      141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy      652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db      161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy      712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db      181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy      772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db      201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy      832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db      221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy      892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db      241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy      952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db      261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy     1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db      281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy     1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db      301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy     1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db      321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340

Qy     1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
Db      341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy     1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
Db      361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy     1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
Db      381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy     1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
Db      401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy     1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
Db      421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy     1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
Db      441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy     1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
Db      461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy     1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
Db      481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy     1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
Db      501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy     1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
Db      521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy     1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
Db      541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy     1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
Db      561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy     1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
Db      581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy     1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
Db      601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGly 620

Qy     2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
Db      621 LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy     2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
Db      641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy     2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
Db      661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSer 680

Qy     2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
Db      681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlnLysValTyrLeu 700

Qy     2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
Db      701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy     2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db      721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 12
US-07-743-357-10
; Sequence 10, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
;   APPLICANT: Kang, Yong C.
;   TITLE OF INVENTION: Polypeptide having immunological
;   TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:

```
;       ADDRESSEE:  KIRBY EADES GALE BAKER
;       STREET:  Box 3432, Station D
;       CITY:  Ottawa
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  K1M 1H8
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/743,357
;       FILING DATE:  21-AUG-1991
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/CA90/00062
;       FILING DATE:  23-FEB-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Gale, Edwin J.
;       REGISTRATION NUMBER:  28,584
;       REFERENCE/DOCKET NUMBER:  30924-2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (613) 237-6900
;       TELEFAX:  (613) 237-0045
;   INFORMATION FOR SEQ ID NO:  10:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1003 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  protein
;     HYPOTHETICAL:  NO
;     FRAGMENT TYPE:  internal
;     ORIGINAL SOURCE:
;       ORGANISM:  Human immunodeficiency virus type 1
;       STRAIN:  ELI
US-07-743-357-10

Alignment Scores:
Pred. No.:              7.76e-211     Length:          1003
Score:                  3497.50       Matches:         650
Percent Similarity:     95.83%        Conservative:    40
Best Local Similarity:  90.28%        Mismatches:      27
Query Match:            76.31%        Indels:          3
DB:                     2             Gaps:            2

US-09-610-313B-30 (1-2469) x US-07-743-357-10 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||:::|||||||||||||||||||||||      |||        :::|||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerProLysGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
                 |||||||||||||||||||||||||||:::|||   |||   ||||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTG 390
              |||:::   ||||||||||||||||||:::         ||||||||||||||||||
Db         41 SerLysThrGlyAlaGluArgGlnGlyThrValSerPheAsnPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||||||||:::||||||:::|||||||||:::||||||||||||||||||
Db         61 TrpGlnArgProLeuValAlaIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||:::||||||||||||||||||||||:::||||||||||||||||||||||
Db        121 IleCysGlyGlnLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              |||||||||:::|||||||||:::||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              |||||||||||||||||||||||||||   ||||||   :::||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrAspMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||::::::|||||||||||||||||||||||||||||||:::||||||||||||||||
Db        201 IleSerArgIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              |||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleSerSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230
              |||||||||||||||||   :::|||||||||:::||||||||||||||||||||||||
Db        321 IleLeuGluProPheArgLysGlnAsnProGluMetValIleTyrGlnTyrMetAspAsp 340

Qy       1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290
              ||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLysLeuArg 360

Qy       1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350
              :::|||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuArgTrpGlyPheThrArgProAspLysLysHisGlnLysGluProPro 380

Qy       1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1410
              ||||||||||||||||||||||||||||||||||||||||||||||   |||:::|||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400

Qy       1411 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1470
              |||||||||||||||||||||||||||||||||   ||||||   :::|||||||||
Db        401 ProGluLysGluSerTrpThrValAsnAspIleGlnAsnLeuValGluArgLeuAsnTrp 420

Qy       1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr 440

Qy       1531 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1590
              ||||||||||::::::::|||||||||||||||||||||||||||||||||||||||
Db        441 LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460

Qy       1591 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1650
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1651 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1710
```

```
Db        481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy       1711 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1770
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520

Qy       1771 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1830
Db        521 GlnLeuAlaGluAlaValGlnArgIleSerThrGluSerIleValIleTrpGlyArgThr 540

Qy       1831 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpAlaGluTyrTrp 560

Qy       1891 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1950
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1951 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2010
Db        581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       2011 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2070
Db        601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2071 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
Db        621 ValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640

Qy       2131 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2190
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2191 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2250
Db        661 GlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIle 680

Qy       2251 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2310
Db        681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db        701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 13
US-07-743-357-8
; Sequence 8, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
;   APPLICANT: Kang, Yong C.
;   TITLE OF INVENTION: Polypeptide having immunological
;   TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: KIRBY EADES GALE BAKER
;   STREET: Box 3432, Station D
;   CITY: Ottawa
;   STATE: Ontario
;   COUNTRY: Canada
;   ZIP: K1M 1H8
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/743,357
;   FILING DATE: 21-AUG-1991
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/CA90/00062
;   FILING DATE: 23-FEB-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Gale, Edwin J.
;   REGISTRATION NUMBER: 28,584
;   REFERENCE/DOCKET NUMBER: 30924-2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (613) 237-6900
;   TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1003 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   FRAGMENT TYPE: internal
;   ORIGINAL SOURCE:
;   ORGANISM: Human immunodeficiency virus type 1
;   STRAIN: RF
US-07-743-357-8

```
Alignment Scores:
Pred. No.:               1.2e-210      Length:         1003
Score:                   3494.50       Matches:        651
Percent Similarity:      95.83%        Conservative:   39
Best Local Similarity:   90.42%        Mismatches:     27
Query Match:             76.25%        Indels:         3
DB:                      2             Gaps:           2

US-09-610-313B-30 (1-2469) x US-07-743-357-8 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluLeuSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSerLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTG 390
Db         41 SerGluAlaGlyGluAspArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
Db         61 TrpGlnArgProIleValThrValLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
Db         81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
Db        121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
Db        141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
Db        181 ThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLys 200
```

```
Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
Db        201 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
Db        221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysGluPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
Db        281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProArgIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230
Db        321 IleLeuGluProPheLysLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy       1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgIleLysIleGluGluLeuArg 360

Qy       1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350
Db        361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy       1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1410
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleValLeu 400

Qy       1411 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1470
Db        401 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy       1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530
Db        421 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 440

Qy       1531 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1590
Db        441 LysAlaLeuThrGluValValGlnLeuThrLysGluAlaGluLeuGluLeuAlaGluAsn 460

Qy       1591 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1650
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1651 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1710
Db        481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy       1711 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1770
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520

Qy       1771 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1830
Db        521 GlnLeuThrGluAlaValGlnLysValAlaThrGluSerIleValIleTrpGlyLysThr 540

Qy       1831 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890
Db        541 ProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy       1891 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1950
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1951 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2010
Db        581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       2011 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2070
Db        601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2071 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
Db        621 ValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy       2131 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2190
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2191 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2250
Db        661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680

Qy       2251 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2310
Db        681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db        701 GlnValAspArgLeuValSerThrGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 14
US-07-743-357-7
; Sequence 7, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1004 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein

```
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human immunodeficiency virus type 1
;      STRAIN:  SF2
US-07-743-357-7

Alignment Scores:
Pred. No.:                3.52e-210      Length:         1004
Score:                    3487.00        Matches:        650
Percent Similarity:       95.42%         Conservative:   38
Best Local Similarity:    90.15%         Mismatches:     29
Query Match:              76.09%         Indels:         4
DB:                       2              Gaps:           2

US-09-610-313B-30 (1-2469) x US-07-743-357-7 (1-1004)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy    280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------GACAACCCC 333
Db     21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40

Qy    334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db     41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy    388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db     61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy    448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db     81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100

Qy    508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db    101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProVal 120

Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
Db    321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy   1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
Db    341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy   1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
Db    361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy   1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
Db    381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMet 400

Qy   1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
Db    401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy   1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
Db    421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy   1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
Db    441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy   1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
Db    461 AsnArgGluIleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeu 480

Qy   1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
Db    481 ValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy   1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
Db    501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy   1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
Db    521 LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys 540

Qy   1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
Db    541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyr 560

Qy   1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
Db    561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy   1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
Db    581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy   2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
Db    601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy   2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
Db    621 ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy   2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
```

```
               ||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
              ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
              ||||||:::||||||||||||||   ||||||||||||||||||||||:::||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720

Qy       2368 GAT 2370
              |||
Db        721 Asp 721
```

RESULT 15
US-09-319-588C-6
; Sequence 6, Application US/09319588C
; Patent No. 6509018
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAUCLERE, Philippe
; APPLICANT: LOUSSERT-AJAKA, Ibtissam
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSSI, Francoise
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/09/319,588C
; CURRENT FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1014
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-09-319-588C-6

```
Alignment Scores:
Pred. No.:              1.43e-193      Length:         1014
Score:                  3221.50        Matches:        587
Percent Similarity:     90.80%         Conservative:   74
Best Local Similarity:  80.63%         Mismatches:     56
Query Match:            70.29%         Indels:         11
DB:                     4              Gaps:           3

US-09-610-313B-30 (1-2469) x US-09-319-588C-6 (1-1014)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||:::|||         |||   :::   |||:::   |||   :::
Db          1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProProAspAsn 20

Qy        280 AAC------CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG--------------- 318
              |||      ||||||:::||||||   :::|||||||||   |||
Db         21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40

Qy        319 ---CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG---------GGCACC 366
                    |||:::          |||   |||   :::|||:::            |||
Db         41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
                 |||||||||||||||||||||||||||||:::::::::::::|||:::|||
Db         61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ::::::::||||||||||||||||||||||||||||||||||||||:::|||||||:::
Db         81 GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluGluLeuGln 100

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              |||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuGluGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArg 120

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              |||||||||   |||   ::::::|||   |||:::||||||:::|||||||||||:::
Db        121 GlnTyrAspAsnIleThrValAspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||:::|||||||||:::|||||||||||||
Db        141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||   |||||||||:::||||||   ||||||
Db        181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACAACACC 846
                 ||||||||||||||||||||||||:::::::|||||||||||||||||||||||||
Db        201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGlu 300

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||:::||||||||||||||||||||||||||    ::::::|||||||||
Db        321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              :::|||||||||||||||||||||||||||||||||||||||||||||:::   ||||||
Db        341 IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              |||      :::|||:::|||||||   |||||||||:::||||||||||||||||||||
Db        361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||:::|||||||||||||:::   ||||||||||||||||||||
Db        401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||:::|||
Db        421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||:::|||||||||:::|||||||||::::::|||      |||||||||
```

```
Db        441 LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||:::|||||||||:::|||||||||
Db        461 AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
              |||||||||   |||:::|||||||||||||||||||||||||||      |||||||||
Db        481 TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThr 500

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
              ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        501 TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
              :::|||||||||||||||:::|||||||||   |||   |||:::|||:::|||   |||
Db        521 SerAlaHisThrAsnAspIleLysGlnLeuValGluValValArgLysValAlaThrGlu 540

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||   |||
Db        541 SerIleValIleTrpGlyLysThrProLysPheArgLeuProValGlnLysGluValTrp 560

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
              |||   ||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        561 GluAlaTrpTrpThrAspHisTrpGlnAlaThrTrpIleProGluTrpGluPheValAsn 580

Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
              ||||||||||||||||||||||||||||||||||||   |||||||||   |||||||||
Db        581 ThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGluProIleSerGlyAlaGlu 600

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||:::|||
Db        601 ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
              |||||||||||||||||||||:::||||||:::   :::|||||||||||||||   |||
Db        621 ThrAspArgGlyArgGlnLysValValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
              ||||||||||||   :::|||||||||:::||||||   :::||||||||||||||||||
Db        641 LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
              ||||||||||||:::|||||||||   :::||||||||||||||||||||||||||||||
Db        661 SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
              :::||||||||||||:::|||||||||||||||:::||||||||||||||||||||||||
Db        681 SerGlnIleIleGluGluLeuIleLysLysGluArgValTyrLeuSerTrpValProAla 700

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
              |||||||||||||||||||||||||||:::|||||||||||||||   ||||||||||||
Db        701 HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720

Qy       2347 GTGCTGTTCCTGGACGGCATCGAT 2370
              :::||||||||||||||||||:::
Db        721 IleLeuPheLeuAspGlyIleGlu 728

Search completed: June  2, 2005, 05:11:22
Job time : 112.083 secs
```

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  2, 2005, 04:23:54 ; Search time 153.54 Seconds
                                          (without alignments)
                                          11117.427 Million cell updates/sec

Title:          US-09-610-313B-30
Perfect score:  4583
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2469

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1465611 seqs, 345679903 residues

Total number of hits satisfying chosen parameters:      2931222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155138_15193/app_query.fasta_1.7893
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09610313_@CGN_1_1_550_@runat_31052005_155138_15193
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_AA:*
                1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
                17:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
                19:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                20:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3608	78.7	998	16	US-10-332-413-4	Sequence 4, Appli
2	3539.5	77.2	850	16	US-10-332-413-10	Sequence 10, Appl
3	3535	77.1	1003	14	US-10-283-847-17	Sequence 17, Appl
4	3513	76.7	1015	17	US-10-634-165-9	Sequence 9, Appli
5	3485	76.0	739	16	US-10-093-953A-26	Sequence 26, Appl
6	3485	76.0	739	16	US-10-093-953A-31	Sequence 31, Appl
7	3468	75.7	1003	16	US-10-325-468-23	Sequence 23, Appl
8	3466	75.6	1003	16	US-10-325-468-35	Sequence 35, Appl
9	3465	75.6	1003	16	US-10-325-468-9	Sequence 9, Appli
10	3465	75.6	1003	16	US-10-325-468-20	Sequence 20, Appl
11	3457	75.4	995	15	US-10-296-734-2	Sequence 2, Appli
12	3443	75.1	995	15	US-10-296-734-1470	Sequence 1470, Ap
13	3289.5	71.8	1006	15	US-10-296-734-1471	Sequence 1471, Ap
14	3221.5	70.3	1014	14	US-10-301-661A-6	Sequence 6, Appli
15	3135.5	68.4	1350	10	US-09-952-060-35	Sequence 35, Appl
16	3135.5	68.4	1350	16	US-10-380-641-35	Sequence 35, Appl
17	3135.5	68.4	1350	17	US-10-636-730-35	Sequence 35, Appl
18	2860.5	62.4	875	10	US-09-952-060-6	Sequence 6, Appli
19	2860.5	62.4	875	15	US-10-168-217A-6	Sequence 6, Appli
20	2860.5	62.4	875	16	US-10-380-641-6	Sequence 6, Appli
21	2860.5	62.4	875	17	US-10-636-730-6	Sequence 6, Appli
22	2855	62.3	850	10	US-09-952-060-2	Sequence 2, Appli
23	2855	62.3	850	15	US-10-168-217A-2	Sequence 2, Appli
24	2855	62.3	850	16	US-10-380-641-2	Sequence 2, Appli
25	2855	62.3	850	17	US-10-636-730-2	Sequence 2, Appli
26	2827	61.7	562	9	US-09-735-487-14	Sequence 14, Appl
27	2827	61.7	562	14	US-10-342-188-14	Sequence 14, Appl
28	2823	61.6	560	15	US-10-399-920-2	Sequence 2, Appli
29	2821	61.6	561	14	US-10-059-271-87	Sequence 87, Appl
30	2815	61.4	562	14	US-10-102-622-10	Sequence 10, Appl
31	2814.5	61.4	875	10	US-09-952-060-8	Sequence 8, Appli
32	2814.5	61.4	875	15	US-10-168-217A-8	Sequence 8, Appli
33	2814.5	61.4	875	16	US-10-380-641-8	Sequence 8, Appli
34	2814.5	61.4	875	17	US-10-636-730-8	Sequence 8, Appli
35	2809	61.3	850	10	US-09-952-060-4	Sequence 4, Appli
36	2809	61.3	850	15	US-10-168-217A-4	Sequence 4, Appli
37	2809	61.3	850	16	US-10-380-641-4	Sequence 4, Appli
38	2809	61.3	850	17	US-10-636-730-4	Sequence 4, Appli
39	2796	61.0	560	14	US-10-205-641-1	Sequence 1, Appli
40	2786.5	60.8	561	14	US-10-102-622-12	Sequence 12, Appl
41	2771	60.5	560	11	US-09-725-652-1	Sequence 1, Appli
42	2740	59.8	546	8	US-08-808-031A-29	Sequence 29, Appl
43	2727	59.5	979	14	US-10-271-181B-118	Sequence 118, App
44	2626	57.3	851	14	US-10-059-271-88	Sequence 88, Appl
45	2609	56.9	999	15	US-10-346-000A-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-10-332-413-4
; Sequence 4, Application US/10332413
; Publication No. US20040116660A1
; GENERAL INFORMATION:
;  APPLICANT: Johnstone, Robert Edward
;  APPLICANT:  Swanstrom, Ronald Ivar
;  APPLICANT:  Morris, Lynn
;  APPLICANT:  Karim, Salim Abdool
;  APPLICANT:  Williamson, Carolyn
;  TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H
;  TITLE OF INVENTION:  Subtype Isolates, Their Genes nd Modifications and Derivatives T
;  FILE REFERENCE: 45669-281993
;  CURRENT APPLICATION NUMBER: US/10/332,413
;  CURRENT FILING DATE:  2003-09-12
;  PRIOR APPLICATION NUMBER: PCT/IB01/01208
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/216,995
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: ZA 2000/3437
;  PRIOR FILING DATE: 2000-07-10
;  PRIOR APPLICATION NUMBER: ZA 2000/4924
```

```
;  PRIOR FILING DATE: 2000-09-15
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 998
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-10-332-413-4

Alignment Scores:
Pred. No.:              3.5e-168        Length:         998
Score:                  3608.00         Matches:        670
Percent Similarity:     97.63%          Conservative:   29
Best Local Similarity:  93.58%          Mismatches:     17
Query Match:            78.73%          Indels:         0
DB:                     16              Gaps:           0

US-09-610-313B-30 (1-2469) x US-10-332-413-4 (1-998)

Qy        223 TTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAAC 282
              ||||||||:::||||||||||||||||:::||||||||||||||||||||||||
Db          1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20

Qy        283 CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAG 342
              ||||||||||||||||||||||||||||||||||||||    :::|||||||||||||||
Db         21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu 40

Qy        343 GCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
                 |||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db         41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro 60

Qy        403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
              ||||||||||||||:::|||||||||    :::||||||||||||||||||||||||||
Db         61 LeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp 80

Qy        463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
              |||||||||||||||::::::::::||||||||||||||||||||||||||||||||||||
Db         81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 100

Qy        523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 IleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLys 120

Qy        583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140

Qy        643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160

Qy        703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180

Qy        763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200

Qy        823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
              |||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        201 GlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220

Qy        883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240

Qy        943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp 260

Qy       1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
              ||||||||||||||||||||||||||||||||||||    ||||||||||||||||||||
Db        261 ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 280

Qy       1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 300

Qy       1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
              ||||||||||||||||||||:::|||||||||    ||||||||||||||||||||||||
Db        301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320

Qy       1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGC 1242
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 340

Qy       1243 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1302
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeu 360

Qy       1303 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATG 1362
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380

Qy       1363 GGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1422
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||:::
Db        381 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400

Qy       1423 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1482
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420

Qy       1483 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1542
              |||||||||||||||||||||||||||||||||||||||||||||    |||||||||||
Db        421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440

Qy       1543 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1602
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 AspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu 460

Qy       1603 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1662
              :::||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        461 LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln 480

Qy       1663 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1722
              |||||||||    |||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 500

Qy       1723 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1782
              ||||||||||||||||||     ||||||    ||||||||||||||||||||||||||
Db        501 ThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThrGlu 520

Qy       1783 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1842
              ||||||||||||||||::::::||||||||||||||||||||||||||||||||||||||
Db        521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540

Qy       1843 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1902
              |||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db        541 LeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThrTrp 560

Qy       1903 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 580

Qy       1963 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2022
              ||||||||||||    ||||||||||||||||||||||||||||||||||||||||||||
Db        581 LysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 600

Qy       2023 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2082
              ||||||||||||||||||||||||||||||||||||||||||||||:::||:::
Db        601 LysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeuSer 620

Qy       2083 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640
```

```
Qy       2143 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2202
Db        641 SerGluValAsnIleIleThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 660

Qy       2203 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2262
Db        661 AspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluArg 680

Qy       2263 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2322
Db        681 ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700

Qy       2323 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db        701 LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716
```

```
RESULT 2
US-10-332-413-10
; Sequence 10, Application US/10332413
; Publication No. US20040116660A1
; GENERAL INFORMATION:
;  APPLICANT: Johnstone, Robert Edward
;  APPLICANT:  Swanstrom, Ronald Ivar
;  APPLICANT:  Morris, Lynn
;  APPLICANT:  Karim, Salim Abdool
;  APPLICANT:  Williamson, Carolyn
;  TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H
;  TITLE OF INVENTION:  Subtype Isolates, Their Genes nd Modifications and Derivatives T
;  FILE REFERENCE: 45669-281993
;  CURRENT APPLICATION NUMBER: US/10/332,413
;  CURRENT FILING DATE:  2003-09-12
;  PRIOR APPLICATION NUMBER: PCT/IB01/01208
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/216,995
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: ZA 2000/3437
;  PRIOR FILING DATE: 2000-07-10
;  PRIOR APPLICATION NUMBER: ZA 2000/4924
;  PRIOR FILING DATE: 2000-09-15
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
;   LENGTH: 850
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-10-332-413-10

Alignment Scores:
Pred. No.:              7.62e-165      Length:         850
Score:                  3539.50        Matches:        669
Percent Similarity:     91.18%         Conservative:   34
Best Local Similarity:  86.77%         Mismatches:     45
Query Match:            77.23%         Indels:         23
DB:                     16             Gaps:           4

US-09-610-313B-30 (1-2469) x US-10-332-413-10 (1-850)

Qy         97 TGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAG 156
Db         84 CysAlaThrValGlyLys------------GlyAspArgCysGlyProLeuArgTyrTyr 99

Qy        157 GGCTGCTGGAAG-----------------------TGCGGCAAGGAGGGCCACCAGATG 192
Db        100 AlaSerTrpArgLysGlyAspValLeuGlnGlyAspValGlyArgGlnGlyPheProSer 119

Qy        193 AAGGACTGCACCGAGCGCCAG------------GCCAACTTCTTCCGCGAGGACCTG--- 237
Db        120 HisAspValValLysArgArgProValProSerLeuHisAlaCysArgSerThrLeuGlu 139

Qy        238 ------------------GCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db        140 AspProArgValProSerSerPheProGlnGlyProAlaArgGlnPheProSerGluGln 159

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
Db        160 ThrArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAspAsnProArgSer 179

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGC 399
Db        180 GluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg 199

Qy        400 CCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCC 459
Db        200 ProLeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAla 219

Qy        460 GACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGC 519
Db        220 AspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGly 239

Qy        520 GGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGC 579
Db        240 GlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGly 259

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
Db        260 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 279

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
Db        280 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 299

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
Db        300 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluVal 319

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
Db        320 LysIleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLys 339

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
Db        340 IleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysGluAspSerThr 359

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
Db        360 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 379

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
Db        380 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 399

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
Db        400 AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAla 419

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
Db        420 PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 439

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
Db        440 ProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSerMetThrLysIleLeuGlu 459

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTG 1239
Db        460 ProPheArgAlaLysAsnProGluIleValIleTyrGlnTyrMetAlaAlaLeuTyrVal 479

Qy       1240 GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTG 1299
Db        480 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeu 499

Qy       1300 CTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGG 1359
Db        500 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 519
```

```
Qy       1360 ATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAG 1419
Db        520 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLys 539

Qy       1420 GAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG 1479
Db        540 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpThrSerGln 559

Qy       1480 ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTG 1539
Db        560 IleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 579

Qy       1540 ACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATC 1599
Db        580 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIle 599

Qy       1600 CTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATC 1659
Db        600 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 619

Qy       1660 CAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTG 1719
Db        620 GlnLysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 639

Qy       1720 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACC 1779
Db        640 LysThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThr 659

Qy       1780 GAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTC 1839
Db        660 GluAlaValGlnLysIleSerLeuGluSerIleValThrTrpGlyLysThrProLysPhe 679

Qy       1840 CGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACC 1899
Db        680 ArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThr 699

Qy       1900 TGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTG 1959
Db        700 TrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeu 719

Qy       1960 GAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAG 2019
Db        720 GluLysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 739

Qy       2020 ACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTG 2079
Db        740 ThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeu 759

Qy       2080 ACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGC 2139
Db        760 SerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 779

Qy       2140 GGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAG 2199
Db        780 GluSerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln 799

Qy       2200 CCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAG 2259
Db        800 ProAspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlu 819

Qy       2260 AAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGAC 2319
Db        820 ArgAlaTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAspGluGlnValAsp 839

Qy       2320 AAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
Db        840 LysLeuValSerSerGlyIleArgLysValLeu 850
```

RESULT 3
US-10-283-847-17
; Sequence 17, Application US/10283847
; Publication No. US20030162720A1
; GENERAL INFORMATION:
; APPLICANT: Nicolette, Charles A.
; APPLICANT: Walker, Bruce
; TITLE OF INVENTION: THERAPEUTIC ANTI-HIV (IV9) COMPOUNDS
; FILE REFERENCE: GZ 2112.00
; CURRENT APPLICATION NUMBER: US/10/283,847
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 60/345,116
; PRIOR FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus
US-10-283-847-17

Alignment Scores:

Pred. No.:	1.28e-164	Length:	1003
Score:	3535.00	Matches:	659
Percent Similarity:	96.12%	Conservative:	34
Best Local Similarity:	91.40%	Mismatches:	24
Query Match:	77.13%	Indels:	4
DB:	14	Gaps:	2

US-09-610-313B-30 (1-2469) x US-10-283-847-17 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db         41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
```

```
Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
              ||||||||||||||||||      :::||||||:::||||||||||||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
              |||||||||||||||||||||||||||||||||||||||      |||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
              |||:::|||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
                  |||||||||||::::::::|||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
              :::||||||||||||||||||||||      |||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
              ||||||||||||||||||||||||||||:::||||||    |||||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
              ||||||||||||||||||||||||||||||        ||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
              |||||||||||||||||:::|||||||||||||||||||:::||||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
              :::|||:::||||||:::|||||||||||||||||||||||||||||||    ||||||
Db        621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
              ||||||||||||||     |||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
              ||||||:::||||||||||||||||      |||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370
              |||
Db        721 Asp 721
```

RESULT 4
US-10-634-165-9
; Sequence 9, Application US/10634165
; Publication No. US20050095581A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Nancy T
; APPLICANT: Gallo, Robert C.
; APPLICANT: Wong-Staal, Flossie
; TITLE OF INVENTION: DETECTION OF HIV-1 DNA
; FILE REFERENCE: 223695
; CURRENT APPLICATION NUMBER: US/10/634,165
; CURRENT FILING DATE: 2003-08-05
; PRIOR APPLICATION NUMBER: 08/463,028
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 06/693,866
; PRIOR FILING DATE: 1985-01-23
; PRIOR APPLICATION NUMBER: 06/659,339
; PRIOR FILING DATE: 1984-10-10
; PRIOR APPLICATION NUMBER: 06/643,306
; PRIOR FILING DATE: 1984-08-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
;   LENGTH: 1015
;   TYPE: PRT
;   ORGANISM: T cell leukemia-lymphoma virus (HTLV)
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1015)
;   OTHER INFORMATION: /note= "pol protein of HTLV-III"
US-10-634-165-9

```
Alignment Scores:
Pred. No.:                1.52e-163     Length:         1015
Score:.                   3513.00       Matches:        658
Percent Similarity:       94.27%        Conservative:   33
Best Local Similarity:    89.77%        Mismatches:     26
Query Match:              76.65%        Indels:         16
DB:                       17            Gaps:           3

US-09-610-313B-30 (1-2469) x US-10-634-165-9 (1-1015)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
              ||||||||||||||||||||||    ||||||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
```

```
Qy        268 -----------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
Db        641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 5

```
US-10-093-953A-26
; Sequence 26, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
;  APPLICANT: Robinson, Harriet L.
;  APPLICANT:  Smith, James M.
;  APPLICANT:  Hua, Jian
;  APPLICANT:  Moss, Bernard
;  APPLICANT:  Amara, Rama
;  APPLICANT:  Wyatt, Linda
;  APPLICANT:  Earl, Patricia
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
;  TITLE OF INVENTION:  AN IMMUNE RESPONSE
;  FILE REFERENCE: 12804-005002
;  CURRENT APPLICATION NUMBER: US/10/093,953A
;  CURRENT FILING DATE:  2002-03-08
;  PRIOR APPLICATION NUMBER: US 60/186,364
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 60/251,083
;  PRIOR FILING DATE: 2000-12-01
;  PRIOR APPLICATION NUMBER: US 09/798,675
;  PRIOR FILING DATE: 2001-03-02
;  PRIOR APPLICATION NUMBER: PCT/US01/06795
;  PRIOR FILING DATE: 2001-03-02
;  PRIOR APPLICATION NUMBER: US 60/324,845
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/325,004
;  PRIOR FILING DATE: 2001-09-26
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
;   LENGTH: 739
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: protein encoded by construct of vaccine vector
;   OTHER INFORMATION: pGA2 and insert JS2 expressing clade HIV-1 VL
US-10-093-953A-26

Alignment Scores:
Pred. No.:              3.45e-162     Length:        739
Score:                  3485.00       Matches:       654
Percent Similarity:     94.00%        Conservative:  35
Best Local Similarity:  89.22%        Mismatches:    28
Query Match:            76.04%        Indels:        16
DB:                     16            Gaps:          3

US-09-610-313B-30 (1-2469) x US-10-093-953A-26 (1-739)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||||||||||||||   ||||||||||||||||||||    ||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGC------------------------------------GCCAACAGCCCCACCAGC 303
                 |||                                     ||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||   :::|||     ||||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      ||||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::||||||||:::||||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              ||||||||||||||||||||||||||||||||||||||||||||||||    ||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::|||||||||||||||||||||||||||||||||:::|||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
              |||      ||||||   |||||||||||||||||||||:::||||||||||||||||||
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||:::
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
              :::||||||:::||||||||||||||||||||:::|||||||||||||||||||||||||
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360

Qy       1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
              ||||||||||||||||||||||||||||||   ||||||||||||:::||||||||||||
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440

Qy       1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
              ||||||||||||||||||||||||||||||||||   ||||||||||||:::::::::
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
```

```
              |||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
              ||||||:::||||||   |||||||||||||||||||   ||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540

Qy       1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
              ||||||      |||||||||||||||||||||||||||||||||:::|||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
              |||:::|||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
              ||||||||||||||||||::::::||||||||||||:::|||   ||||||   ||||||
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
              ||||||||||||:::|||||||||||   ||||||||||||||||||||   ||||||
Db        641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
              :::||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
                 ||||||||||:::||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 6
US-10-093-953A-31
; Sequence 31, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; APPLICANT: Moss, Bernard
; APPLICANT: Amara, Rama
; APPLICANT: Wyatt, Linda
; APPLICANT: Earl, Patricia
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; TITLE OF INVENTION: AN IMMUNE RESPONSE
; FILE REFERENCE: 12804-005002
; CURRENT APPLICATION NUMBER: US/10/093,953A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/798,675
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/325,004
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 739
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: protein encoded by construct of vaccine vector
;   OTHER INFORMATION: pGA1 and vaccine insert expressing clade B HIV-1
;   OTHER INFORMATION: Gag-Pol
US-10-093-953A-31

```
Alignment Scores:
Pred. No.:                3.45e-162     Length:          739
Score:                    3485.00       Matches:         654
Percent Similarity:       94.00%        Conservative:    35
Best Local Similarity:    89.22%        Mismatches:      28
Query Match:              76.04%        Indels:          16
DB:                       16            Gaps:            3

US-09-610-313B-30 (1-2469) x US-10-093-953A-31 (1-739)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||   |||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGC-----------------------------------GCCAACAGCCCCACCAGC 303
                 |||                                   ||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||   :::|||   ||||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      |||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::||||||||||||||||||||||||||||||||||:::||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
              |||      ||||||   ||||||||||||||||||||||:::||||||||||||||||
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
```

```
Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||:::
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
              :::||||||:::|||||||||||||||||||:::||||||||||||||||||||||||||
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360

Qy       1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
              |||||||||||||||||||||||||||||||||   |||||||||||:::||||||||||
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440

Qy       1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
              |||||||||||||||||||||||||||||||||||   |||||||||||:::::::::
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
                       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
              ||||||:::||||||   |||||||||||||||||||||   ||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540

Qy       1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
              ||||||         ||||||||||||||||||||||||||||||:::|||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
              |||:::||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
              ||||||||||||||||:::::|||||||||||:::|||   ||||||   ||||||
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
              ||||||||||:::|||||||||||   ||||||||||||||||||||   ||||||
Db        641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
              :::|||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
                 ||||||||||:::||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 7
US-10-325-468-23
; Sequence 23, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone 1.10 protein Pol
US-10-325-468-23

```
Alignment Scores:
Pred. No.:              2.38e-161     Length:        1003
Score:                  3468.00       Matches:       642
Percent Similarity:     95.28%        Conservative:  45
Best Local Similarity:  89.04%        Mismatches:    30
Query Match:            75.67%        Indels:        4
DB:                     16            Gaps:          2

US-09-610-313B-30 (1-2469) x US-10-325-468-23 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
              ||||||||||||||      ||||||   ||||||      |||   :::|||
Db         21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
```

```
Qy    334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db     41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy    388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db     61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy    448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db     81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy    508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db    101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
Db    321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy   1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
Db    341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy   1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
Db    361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy   1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
Db    381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy   1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
Db    401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy   1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
Db    421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy   1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
Db    441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy   1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
Db    461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy   1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
Db    481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy   1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
Db    501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgSerAlaHisThrAsnAspVal 520

Qy   1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
Db    521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy   1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
Db    541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy   1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
Db    561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy   1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
Db    581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy   2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
Db    601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy   2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
Db    621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy   2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
Db    641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy   2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
Db    661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy   2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
Db    681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy   2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
Db    701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy   2368 GAT 2370
Db    721 Glu 721
```

RESULT 8
US-10-325-468-35
; Sequence 35, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:

```
;  APPLICANT: Soong, Nay Wei
;  APPLICANT:  Pekrun, Katja
;  APPLICANT:  Shibata, Riri
;  TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
;  TITLE OF INVENTION:  MODELS OF HIV-1 PATHOGENESIS
;  FILE REFERENCE: 0166.210US
;  CURRENT APPLICATION NUMBER: US/10/325,468
;  CURRENT FILING DATE:  2002-12-19
;  PRIOR APPLICATION NUMBER: US 60/343,524
;  PRIOR FILING DATE: 2001-12-21
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone P8A26 protein Pol
US-10-325-468-35

Alignment Scores:
Pred. No.:              2.98e-161      Length:          1003
Score:                  3466.00        Matches:         643
Percent Similarity:     95.28%         Conservative:    44
Best Local Similarity:  89.18%         Mismatches:      30
Query Match:            75.63%         Indels:          4
DB:                     16             Gaps:            3

US-09-610-313B-30 (1-2469) x US-10-325-468-35 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGAC---AACCCC 333
                 ||||||||||||       ||||||   ||||||   ||||||||     |||
Db         21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgGlyAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||:::|||||||||||:::        ||||||||||||||||
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              ||||||||||||||||||||||||:::||||||:::|||||||||:::||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              |||||||||||||||||||||||||||||:::|||   ||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              |||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              :::||||||||   ||||||:::||||||||||||:::||||||||||||||||||||||
Db        121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||||:::||||||||:::||||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              |||||||||||||||||||||||||       ||||||   ||||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||::::::||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
              |||   ||||||||||||||     :::||||||:::|||:::||||||||||||||||||
Db        321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
              |||:::||||||||::::||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
              ||||||||||||||||||     |||||||||||||:::|||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
                 ||||||||||:::::::|||||||||||||:::||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
              :::   ||||||||||||||||||         ||||||||||||||||||:::||||||||
Db        481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
              ||||||||||||||||||||||||||||||||||||   |||    |||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
                 |||||||||:::|||||||||||||||||||||||||||||||||||||||:::|||
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
```

```
Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy       2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
Db        681 IleLysLysGluLysValTyrLeuThrTrpIleProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370
Db        721 Glu 721
```

RESULT 9
US-10-325-468-9
; Sequence 9, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone 1.4 protein Pol;
;   OTHER INFORMATION: clone 1.26 protein Pol; clone P10.21 protein Pol;
;   OTHER INFORMATION: clone P10.26 protein Pol
US-10-325-468-9

Alignment Scores:

Pred. No.:	3.33e-161	Length:	1003
Score:	3465.00	Matches:	642
Percent Similarity:	95.15%	Conservative:	44
Best Local Similarity:	89.04%	Mismatches:	31
Query Match:	75.61%	Indels:	4
DB:	16	Gaps:	2

US-09-610-313B-30 (1-2469) x US-10-325-468-9 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db         21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db         81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
Db        321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
Db        361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
              |||||||||||||||||   |||||||||||:::||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
                 ||||||||||:::::::||||||||:::||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
              :::   |||||||||||||||||        ||||||||||||||||:::|||||||||
Db        481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
              ||||||||||||||||||||||||||||||||   |||   ||||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
              ||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
                 |||||||||:::|||||||||||||||||||||||||||||||||||||:::|||
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
              ||||||||||||||||||:::||||||||||||||||||||   |||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy       2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
              :::||||||||||||:::|||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
              ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
              ||||||:::||||||||||||||   |||||||||:::||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370
              :::
Db        721 Glu 721
```

RESULT 10
US-10-325-468-20

```
; Sequence 20, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
;  APPLICANT: Soong, Nay Wei
;  APPLICANT:  Pekrun, Katja
;  APPLICANT:  Shibata, Riri
;  TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
;  TITLE OF INVENTION:  MODELS OF HIV-1 PATHOGENESIS
;  FILE REFERENCE: 0166.210US
;  CURRENT APPLICATION NUMBER: US/10/325,468
;  CURRENT FILING DATE:  2002-12-19
;  PRIOR APPLICATION NUMBER: US 60/343,524
;  PRIOR FILING DATE: 2001-12-21
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone 1.27 protein Pol
US-10-325-468-20

Alignment Scores:
Pred. No.:               3.33e-161      Length:        1003
Score:                   3465.00        Matches:       642
Percent Similarity:      95.15%         Conservative:  44
Best Local Similarity:   89.04%         Mismatches:    31
Query Match:             75.61%         Indels:        4
DB:                      16             Gaps:          2

US-09-610-313B-30 (1-2469) x US-10-325-468-20 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                 ||||||||||||||      ||||||   ||||||      |||   :::|||
Db         21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||||:::||||||||||||:::      |||||||||||||||||
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              ||||||||||||||||||||:::||||||:::||||||||||:::||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              |||||||||||||||||||||||||||||:::|||   ||||||||||:::||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              :::|||||||||   ||||||:::|||||||||||||:::||||||||||||||||||||
Db        121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||||      ||||||   ||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
```

```
Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
Db        321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
Db        361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
Db        481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy       2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370
Db        721 Glu 721
```

RESULT 11
US-10-296-734-2
; Sequence 2, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 995
;   TYPE: PRT
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: POL consensus polypeptide
US-10-296-734-2

Alignment Scores:

Pred. No.:	8.19e-161	Length:	995
Score:	3457.00	Matches:	644
Percent Similarity:	95.00%	Conservative:	40
Best Local Similarity:	89.44%	Mismatches:	26
Query Match:	75.43%	Indels:	10
DB:	15	Gaps:	3

US-09-610-313B-30 (1-2469) x US-10-296-734-2 (1-995)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
```

```
                 |||||||||    ||||||:::|||        ||||||
Db          21 ThrGlyAlaAsnSerSerAlaSerArgLysLeu---------GlyAspGly--------- 34

Qy         340 GAGGCCGGCGCCGAGCGCCAGGGCACC---------CTGAACTTCCCCAGATCACCCTG 390
                  ||||||||||||||||||||||||             :::||||||||||||||||||
Db          35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheProGlnIleThrLeu 53

Qy         391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
               |||||||||||||||||:::||||||:::|||||||||:::||||||||||||||||||
Db          54 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 73

Qy         451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
               ||||||||||||||||||||||||||:::::::::|||||||||||||||||||||||||
Db          74 ThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLys 93

Qy         511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          94 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 113

Qy         571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
               ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db         114 IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133

Qy         631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
               |||||||||||||||||||:::||||||||||||||||||||||||||||||||||:::
Db         134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153

Qy         691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 173

Qy         751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
               |||||||||||||||||||||||||      ||||||:::|||||||||:::||||||||
Db         174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGluGlyLys 193

Qy         811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
               |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         194 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 213

Qy         871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         214 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233

Qy         931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 253

Qy         991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
               ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db         254 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLys 273

Qy        1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 293

Qy        1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
               ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db         294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313

Qy        1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230
               |||||||||||||||||      :::||||||||:::||||||||||||||||||||||
Db         314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTyrGlnTyrMetAspAsp 333

Qy        1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290
               ||||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db         334 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353

Qy        1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         354 AlaHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373

Qy        1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1410
               ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db         374 PheLeuTrpMetGlyTyrGluLeuHisProAspArgTrpThrValGlnProIleGluLeu 393

Qy        1411 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1470
               |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413

Qy        1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530
               |||||||||||||||   ||||||||||||:::|||||||||||||||||||||||||
Db         414 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 433

Qy        1531 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1590
               ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db         434 LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuGluGluAsn 453

Qy        1591 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1650
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         454 ArgGluIleLeuArgGluProValHisGlyValTyrTyrAspProSerLysAspLeuVal 473

Qy        1651 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1710
               ||||||:::||||||||||||     ||||||||||||||||||||||||||||||||||
Db         474 AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPhe 493

Qy        1711 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1770
               ||||||||||||||||||||||||::::::     |||:::|||||||||||||||||:::
Db         494 LysAsnLeuLysThrGlyLysTyrSerArgLysArgSerAlaHisThrAsnAspValArg 513

Qy        1771 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1830
               ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db         514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533

Qy        1831 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890
               ||||||||||||||||||||||:::|||||||||||||||||||||      :::||||||
Db         534 ProLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTyrTrp 553

Qy        1891 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1950
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 573

Qy        1951 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2010
               ||||||||||||||:::||||||:::||||||||||||||||||||||||||||||||||
Db         574 TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 593

Qy        2011 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2070
               :::|||||||||||||:::||||||||||||||||||||||||||||||||||||||:::
Db         594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613

Qy        2071 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
               :::||||||||||||||||||||||||||||||||||||    ||||||    ||||||||
Db         614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633

Qy        2131 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2190
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653

Qy        2191 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2250
               ||||||||||||||:::||||||||||||:::|||:::||||||||||||:::||||||
Db         654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673

Qy        2251 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2310
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693

Qy        2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
               |||:::|||||||||||||      ||||||||||||||||||||||||||||||||:::
Db         694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713
```

RESULT 12
US-10-296-734-1470
; Sequence 1470, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A

```
;  TITLE OF INVENTION: Synthetic molecules and uses therefor
;  FILE REFERENCE: Savine
;  CURRENT APPLICATION NUMBER: US/10/296,734
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: AU PQ7761/00
;  PRIOR FILING DATE: 2000-05-26
;  NUMBER OF SEQ ID NOS: 1507
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1470
;   LENGTH: 995
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus
US-10-296-734-1470

Alignment Scores:
Pred. No.:               3.95e-160      Length:        995
Score:                   3443.00        Matches:       642
Percent Similarity:      94.72%         Conservative:  40
Best Local Similarity:   89.17%         Mismatches:    28
Query Match:             75.13%         Indels:        10
DB:                      15             Gaps:          3

US-09-610-313B-30 (1-2469) x US-10-296-734-1470 (1-995)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||:::|||||||   |||||||||||||||||   |||||||||
Db          1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
                    ||||||||      ||||||:::|||         ||||||
Db         21 ThrGlyAlaAsnSerSerAlaSerArgLysLeu---------GlyAspGly--------- 34

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACC---------CTGAACTTCCCCCAGATCACCCTG 390
                    ||||||||||||||||||||            :::||||||||||||||||
Db         35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheProGlnIleThrLeu 53

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||||||||:::||||||:::|||||||||:::|||||||||||||||||
Db         54 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 73

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              ||||||||||||||||||||||||||||||:::::::::||||||||||||||||||||||
Db         74 ThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLys 93

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         94 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 113

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        114 IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||:::
Db        134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 173

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              |||||||||||||||||||||||||     ||||||:::|||||||||:::|||||||||
Db        174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGluGlyLys 193

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 213

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 253

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        254 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLys 273

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 293

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db        294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230
              ||||||||||||||||||    :::|||||||||:::|||||||||||||||||||||||
Db        314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTyrGlnTyrMetAspAsp 333

Qy       1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db        334 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353

Qy       1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350
                    |||||||||      ||||||||||||||||||||||||||||||||||||||||
Db        354 AlaHisLeuLeuSerTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373

Qy       1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1410
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db        374 PheLeuTrpMetGlyTyrGluLeuHisProAspArgTrpThrValGlnProIleGluLeu 393

Qy       1411 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1470
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413

Qy       1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530
              ||||||||||||||     ||||||||||||:::|||||||||||||||||||||||
Db        414 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 433

Qy       1531 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1590
              |||||||||||||||||||||||||||||||||||||||||||||||||     ||||||
Db        434 LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuGluGluAsn 453

Qy       1591 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1650
              ||||||||||||||     |||||||||||||||||||||||||||||||||||||||||
Db        454 ArgGluIleLeuArgIleProValHisGlyValTyrTyrAspProSerLysAspLeuVal 473

Qy       1651 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1710
              ||||||:::|||||||||||    ||||||||||||||||||||||||||||||||||||
Db        474 AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPhe 493

Qy       1711 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1770
              ||||||||||||||||||||||::::::      |||:::|||||||||||||||||:::
Db        494 LysAsnLeuLysThrGlyLysTyrSerArgLysArgSerAlaHisThrAsnAspValArg 513

Qy       1771 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1830
              |||||||||||     ||||||||||||||||||||||||||||||||||||||||||||
Db        514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533

Qy       1831 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890
              ||||||||||||||||||||||||:::||||||||||||||||||||     :::||||||
Db        534 ProLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTyrTrp 553

Qy       1891 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 573

Qy       1951 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2010
              |||||||||||||||:::||||||:::|||||||||||||||||||||||||||||||||
Db        574 TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 593

Qy       2011 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2070
              :::||||||||||||:::||||||||||||||||||||||||||||||||||||||||:::
```

```
Db        594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613

Qy       2071 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
              :::||||||||||||||||||||||||||||||||||||    ||||||    ||||||||
Db        614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633

Qy       2131 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653

Qy       2191 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2250
              ||||||||||||:::||||||||||:::|||:::||||||||||||:::|||||||
Db        654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673

Qy       2251 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693

Qy       2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              |||:::||||||||||||         |||||||||||||||||||||||||||||:::
Db        694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713
```

RESULT 13
US-10-296-734-1471
; Sequence 1471, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1471
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: POL Consensus A
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (15)..(15)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (22)..(23)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (26)..(26)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (30)..(30)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (36)..(36)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (38)..(39)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (41)..(44)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (46)..(46)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (48)..(48)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (51)..(51)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (53)..(55)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (58)..(58)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (73)..(73)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (75)..(75)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (105)..(105)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (130)..(130)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (173)..(173)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (194)..(194)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (197)..(197)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (255)..(255)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (280)..(281)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (293)..(293)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (300)..(300)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (316)..(316)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (320)..(320)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (331)..(333)

```
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (335)..(335)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (365)..(365)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (369)..(369)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (373)..(373)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (402)..(403)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (434)..(434)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (439)..(439)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (452)..(452)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (470)..(470)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (480)..(480)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (487)..(487)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (514)..(514)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (534)..(535)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (544)..(544)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (553)..(553)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (558)..(558)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (593)..(593)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (610)..(610)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (675)..(675)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:

Alignment Scores:
Pred. No.:               1.23e-152     Length:         1006
Score:                   3289.50       Matches:        622
Percent Similarity:      89.36%        Conservative:   25
Best Local Similarity:   85.91%        Mismatches:     70
Query Match:             71.78%        Indels:         7
DB:                      15            Gaps:           2

US-09-610-313B-30 (1-2469) x US-10-296-734-1471 (1-1006)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||:::||||||||   ||||||:::||||||   |||   ||||||||
Db          1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArg***PheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCC------ 333
                       ||||||   |||||||||   |||         |||    :::
Db         21 Thr******AsnSer***ThrSerArg***LeuTrpAspGlyGly***Asp******Leu 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGC--------------ACCCTGAACTTCCCC 378
                          |||   |||   |||||                 :::    |||||
Db         41 ************Gly***Glu***GlnGly***Gly*********SerPhe***PhePro 60

Qy        379 CAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAG 438
              ||||||||||||||||||||||||||||||:::::::   :::   ||||||:::   |||
Db         61 GlnIleThrLeuTrpGlnArgProLeuValThrVal***Ile***GlyGlnLeuIleGlu 80

Qy        439 GCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAG 498
              ||||||||||||||||||||||||||||||||||||:::::::::||||||||||||
Db         81 AlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLys 100

Qy        499 TGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAG 558
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        101 TrpLysProLys***IleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGln 120

Qy        559 ATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCC 618
              |||||||||||||||||||||||||||   |||||||||||||||:::||||||||||||
Db        121 IleLeuIleGluIleCysGlyLysLys***IleGlyThrValLeuValGlyProThrPro 140

Qy        619 GTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATC 678
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        141 ValAsnIleIleGlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIle 160

Qy        679 AGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAG 738
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        161 SerProIleGluThrValProValLysLeuLysPro***MetAspGlyProLysValLys 180

Qy        739 CAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAG 798
              ||||||||||||||||||||||||||||||||||||||||   ||||||   |||||||||
Db        181 GlnTrpProLeuThrGluGluLysIleLysAlaLeuThr***IleCys***GluMetGlu 200

Qy        799 AAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCC 858
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAla 220

Qy        859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 IleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys 240

Qy        919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAG 978
              |||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        241 ArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis***AlaGlyLeuLysLys 260

Qy        979 AAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAG 1038
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 LysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAsp*** 280

Qy       1039 GACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATC 1098
```

```
                 |||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db        281 ***PheArgLysTyrThrAlaPheThrIleProSer***AsnAsnGluThrProGly*** 300

Qy       1099 CGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGC 1158
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        301 ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro***IlePheGln*** 320

Qy       1159 AGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG 1218
              ||||||||||||||||||||||||||||||             |||   |||||||||||||||
Db        321 SerMetThrLysIleLeuGluProPheArg*********Pro***IleValIleTyrGln 340

Qy       1219 TACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATC 1278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 TyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIle 360

Qy       1279 GAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAG 1338
              ||||||||||||   |||||||||   ||||||||||   ||||||||||||||||||||
Db        361 GluGluLeuArg***HisLeuLeu***TrpGlyPhe***ThrProAspLysLysHisGln 380

Qy       1339 AAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAG 1398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGln 400

Qy       1399 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1458
              |||         ||||||||||:::|||||||||||||||||||||||||||||||||||
Db        401 Pro******LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 420

Qy       1459 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1518
              ||||||||||||||||||||||||||   |||||||||   :::|||||||||   |||
Db        421 LysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys***LysGlnLeuCys***Leu 440

Qy       1519 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1578
              ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        441 LeuArgGlyAlaLysAlaLeuThrAspIleVal***LeuThrGluGluAlaGluLeuGlu 460

Qy       1579 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1638
              ||||||||||||||||||||||||:::   ||||||||||||||||||||||||||
Db        461 LeuAlaGluAsnArgGluIleLeuLys***ProValHisGlyValTyrTyrAspPro*** 480

Qy       1639 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1698
              ||||||||||||||||||   ||||||||||||   ||||||||||||||||||||||||
Db        481 LysAspLeuValAlaGlu***GlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyr 500

Qy       1699 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1758
              |||||||||||||||||||||||||||||||||||||||      |||:::|||||||||
Db        501 GlnGluProPheLysAsnLeuLysThrGlyLysTyrAla***LysArgSerAlaHisThr 520

Qy       1759 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1818
              ||||||||||||||||||||||||   ||||||||||:::      |||||||||||||||
Db        521 AsnAspValLysGlnLeuThrGluValValGlnLysVal******GluSerIleValIle 540

Qy       1819 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1878
              |||||||||   ||||||||||||||||||||||||   ||||||||||||   ||||||
Db        541 TrpGlyLys***ProLysPheArgLeuProIleGln***GluThrTrpGlu***TrpTrp 560

Qy       1879 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1938
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 MetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu 580

Qy       1939 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1998
              ||||||||||||||||||||||||||:::|||||||   |||||||||||||||||||||
Db        581 ValLysLeuTrpTyrGlnLeuGluLysAspProIle***GlyAlaGluThrPheTyrVal 600

Qy       1999 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2058
              ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        601 AspGlyAlaAlaAsnArgGluThrLys***GlyLysAlaGlyTyrValThrAspArgGly 620

Qy       2059 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2118
              |||||||||:::||||||||||||||||||||||||||||||||||||||||   ||||||
Db        621 ArgGlnLysValValSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIle 640

Qy       2119 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 HisLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAla 660

Qy       2179 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2238
              |||||||||||||||||||||||||:::|||||||||||||   |||||||||||||||
Db        661 LeuGlyIleIleGlnAlaGlnProAspArgSerGluSerGlu***ValAsnGlnIleIle 680

Qy       2239 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2298
              |||:::||||||   |||   ||||||||||||||||||||||||||||||||||||||||
Db        681 GluLysLeuIle***Lys***LysValTyrLeuSerTrpValProAlaHisLysGlyIle 700

Qy       2299 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2358
              ||||||||||||:::||||||||||||||||   ||||||||||||||||||||||||||
Db        701 GlyGlyAsnGluGlnValAspLysLeuValSer***GlyIleArgLysValLeuPheLeu 720

Qy       2359 GACGGCATCGAT 2370
              ||||||||||||
Db        721 AspGlyIleAsp 724
```

RESULT 14

```
US-10-301-661A-6
; Sequence 6, Application US/10301661A
; Publication No. US20030157660A1
; GENERAL INFORMATION:
;  APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-
;  APPLICANT:  INSERM
;  APPLICANT:  ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
;  APPLICANT:  INSTITUT PASTEUR
;  APPLICANT:  MAUCLERE, Philippe
;  APPLICANT:  LOUSSERT-AJAKA, Ibtissam
;  APPLICANT:  SIMON, Francois
;  APPLICANT:  SARAGOSTI, Sentob
;  APPLICANT:  BARRE-SINOUSSI, Francoise
;  TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
;  FILE REFERENCE: 598US12
;  CURRENT APPLICATION NUMBER: US/10/301,661A
;  CURRENT FILING DATE:  2002-11-22
;  PRIOR APPLICATION NUMBER: US/09/319,588C
;  PRIOR FILING DATE: 1999-08-27
;  PRIOR APPLICATION NUMBER: FR96/15087
;  PRIOR FILING DATE: 1996-12-09
;  NUMBER OF SEQ ID NOS: 98
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1014
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-10-301-661A-6

Alignment Scores:
Pred. No.:               2.56e-149      Length:        1014
Score:                   3221.50        Matches:       587
Percent Similarity:      90.80%         Conservative:  74
Best Local Similarity:   80.63%         Mismatches:    56
Query Match:             70.29%         Indels:        11
DB:                      14             Gaps:          3

US-09-610-313B-30 (1-2469) x US-10-301-661A-6 (1-1014)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||:::|||         |||   :::   |||:::   |||   :::
Db          1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProProAspAsn 20

Qy        280 AAC------CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG-------------- 318
              |||      ||||||:::||||||   :::|||||||||   |||
Db         21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40

Qy        319 ---CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG---------GGCACC 366
                    |||:::            |||   |||   :::|||:::            |||
Db         41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
                 |||||||||||||||||||||||||||||||:::::::::::|||:::|||
Db         61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80
```

```
Qy    427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db     81 GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluGluLeuGln 100

Qy    487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db    101 LeuGluGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArg 120

Qy    547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db    121 GlnTyrAspAsnIleThrValAspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140

Qy    607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db    141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160

Qy    667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db    161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180

Qy    727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db    181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200

Qy    787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db    201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220

Qy    847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db    221 ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240

Qy    907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db    241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260

Qy    967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db    261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280

Qy   1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db    281 ProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGlu 300

Qy   1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db    301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320

Qy   1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db    321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340

Qy   1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
Db    341 IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360

Qy   1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
Db    361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380

Qy   1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
Db    381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400

Qy   1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
Db    401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420

Qy   1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
Db    421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440
```

```
Qy   1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
Db    441 LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460

Qy   1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
Db    461 AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480

Qy   1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
Db    481 TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThr 500

Qy   1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
Db    501 TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520

Qy   1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
Db    521 SerAlaHisThrAsnAspIleLysGlnLeuValGluValValArgLysValAlaThrGlu 540

Qy   1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
Db    541 SerIleValIleTrpGlyLysThrProLysPheArgLeuProValGlnLysGluValTrp 560

Qy   1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
Db    561 GluAlaTrpTrpThrAspHisTrpGlnAlaThrTrpIleProGluTrpGluPheValAsn 580

Qy   1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
Db    581 ThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGluProIleSerGlyAlaGlu 600

Qy   1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
Db    601 ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620

Qy   2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
Db    621 ThrAspArgGlyArgGlnLysValValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640

Qy   2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
Db    641 LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660

Qy   2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
Db    661 SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680

Qy   2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
Db    681 SerGlnIleIleGluGluLeuIleLysLysGluArgValTyrLeuSerTrpValProAla 700

Qy   2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
Db    701 HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720

Qy   2347 GTGCTGTTCCTGGACGGCATCGAT 2370
Db    721 IleLeuPheLeuAspGlyIleGlu 728
```

```
RESULT 15
US-09-952-060-35
; Sequence 35, Application US/09952060
; Publication No. US20030044421A1
; GENERAL INFORMATION:
;  APPLICANT: Emini, Emilio A.
;  APPLICANT:  Youil, Rima
;  APPLICANT:  Bett, Andrew J.
;  APPLICANT:  Chen, Ling
;  APPLICANT:  Kaslow, David C.
;  APPLICANT:  Shiver, John W.
;  APPLICANT:  Toner, Timothy J.
;  APPLICANT:  Casimiro, Danilo R.
```

```
;  TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
;  TITLE OF INVENTION:  VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
;  TITLE OF INVENTION:  MODIFICATIONS
;  FILE REFERENCE: 20747Y
;  CURRENT APPLICATION NUMBER: US/09/952,060
;  CURRENT FILING DATE:  2001-09-14
;  PRIOR APPLICATION NUMBER: PCT/US01/28861
;  PRIOR FILING DATE: 2001-09-14
;  PRIOR APPLICATION NUMBER: 60/317,814
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: 60/279,056
;  PRIOR FILING DATE: 2001-03-27
;  PRIOR APPLICATION NUMBER: 60/233,180
;  PRIOR FILING DATE: 2000-09-15
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 1350
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Codon optimized gag-IA pol fusion
US-09-952-060-35

Alignment Scores:
Pred. No.:               4.11e-145     Length:         1350
Score:                   3135.50       Matches:        600
Percent Similarity:      81.29%        Conservative:   43
Best Local Similarity:   75.85%        Mismatches:     60
Query Match:             68.42%        Indels:         88
DB:                      10            Gaps:           5

US-09-610-313B-30 (1-2469) x US-09-952-060-35 (1-1350)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACC---AGCGCCAACATCCTGATGCAGCGCGCAGCAAC 69
                 :::|||||||||||||||||    |||   ||||||   |||:::||||||||    |||
Db        363 LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn 382

Qy         70 TTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCC 129
              |||:::      ::::::    :::|||||||||||||||||||    |||||||||||
Db        383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysValGlyHisIleAla 402

Qy        130 CGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAG 189
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422

Qy        190 ATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAG 249
              |||||||||||    |||||||||||||||||            :::   ||||||||   ::
Db        423 MetLysAspCysAsnGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProSerHisLy 442

Qy        250 GGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAG 309
              :   |||||    :::            |||||     ||||||||||||||      ||
Db        442 sGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProGluGluSe 462

Qy        310 CTGCAGGTGCGCG------GCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGC 363
              |     |||               ||||||||:::     :::      |||        :::
Db        462 rPheArgPheGlyGluGluLysThrThrProSerGlnLysGlnGluProIleAspLys-- 481

Qy        364 ACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGC 423
                 |||                                  ||||||    |||::::::
Db        482 GluLeu--------------------------TyrProLeuAlaSerLeuArg------ 490

Qy        424 GGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATG 483

Db        490 ------------------------------------------------------------ 490

Qy        484 AGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTG 543

Db        490 ------------------------------------------------------------ 490

Qy        544 CGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTG 603

Db        490 ------------------------------------------------------------ 490

Qy        604 ATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACC 663
                                 ::::::    |||    :::         :::|||:::
Db        491 ------------------SerLeuPheGlyAsnAspProSerSerGlnMetAla------ 502

Qy        664 CTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGAC 723
                       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 ---------ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAsp 519

Qy        724 GGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATC 783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        520 GlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIle 539

Qy        784 TGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAAC 843
              |||   ||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        540 CysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsn 559

Qy        844 ACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTC 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPhe 579

Qy        904 CGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCC 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPro 599

Qy        964 GCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGC 1023
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db        600 AlaGlyLeuLysLysLysLysSerValThrValLeuAlaValGlyAspAlaTyrPheSer 619

Qy       1024 GTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAAC 1083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 ValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsn 639

Qy       1084 GAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCC 1143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 GluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro 659

Qy       1144 AGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAG 1203
              :::||||||||||||||||||||||||||||||||||||||||      :::||||||:::
Db        660 AlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAsp 679

Qy       1204 ATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAG 1263
              |||||||||||||||||||||         ||||||||||||||||||||||||||||||
Db        680 IleValIleTyrGlnTyrMetAlaAlaLeuTyrValGlySerAspLeuGluIleGlyGln 699

Qy       1264 CACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCC 1323
              ||||||    ||||||||||||||||||:::||||||||||||||||||      ||||||||
Db        700 HisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrPro 719

Qy       1324 GACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGAC 1383
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 AspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAsp 739

Qy       1384 AAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATC 1443
              |||||||||||||||||||||    ||||||||||||:::||||||||||||||||||||
Db        740 LysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIle 759

Qy       1444 CAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGC 1503
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArg 779

Qy       1504 CAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAG 1563
              ||||||||||||||||||||||||    |||||||||||:::::::::||||||||||||
Db        780 GlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGlu 799

Qy       1564 GAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTG 1623
              ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        800 GluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyVal 819

Qy       1624 TACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGG 1683
              ||||||||||||||||||||||:::||||||||||||||||||||||||      ||||||
Db        820 TyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrp 839
```

```
Qy       1684 ACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATG 1743
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        840 ThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859

Qy       1744 CGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATG 1803
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||      
Db        860 ArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr 879

Qy       1804 GAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACC 1863
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        880 GluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGluThr 899

Qy       1864 TGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTG 1923
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        900 TrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVal 919

Qy       1924 AACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCC 1983
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        920 AsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAla 939

Qy       1984 GAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTAC 2043
              |||||||||||||||   ||||||||||||||||||||||||:::|||||||||||||||
Db        940 GluThrPheTyrValAlaGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyr 959

Qy       2044 GTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACC 2103
              ||||||:::|||||||||||||||||||:::||||:::||||||:::|||||||||||||||||||
Db        960 ValThrAsnArgGlyArgGlnLysValValThrLeuThrAspThrThrAsnGlnLysThr 979

Qy       2104 GAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACC 2163
                 ||||||||||||   |||||||||||||||||||||   ||||||||||||||||||
Db        980 AlaLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThr 999

Qy       2164 GACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTG 2223
                 |||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db       1000 AlaSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeu 1019

Qy       2224 GTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCC 2283
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       1020 ValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValPro 1039

Qy       2284 GCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGC 2343
              ||||||||||||||||||||||||||||||:::||||||||||||||||   |||||||||
Db       1040 AlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArg 1059

Qy       2344 AAGGTGCTGTTCCTGGACGGCATCGAT 2370
              |||||||||||||||||||||||||||
Db       1060 LysValLeuPheLeuAspGlyIleAsp 1068
```

Search completed: June 2, 2005, 05:27:02
Job time : 245.54 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  1, 2005, 20:20:55 ; Search time 209.175 Seconds
                                          (without alignments)
                                          12088.668 Million cell updates/sec

Title:          US-09-610-313B-30
Perfect score:  4583
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2469

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters:      3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155136_15147/app_query.fasta_1.7893
-DB=UniProt_03 -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_725_@runat_31052005_155136_15147 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      UniProt_03:*
                1:  uniprot_sprot:*
                2:  uniprot_trembl:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4178	91.2	1427	2	Q9WF92	Q9wf92 human immun
2	4174	91.1	1427	2	Q9WF89	Q9wf89 human immun
3	4170	91.0	1428	2	Q9WF86	Q9wf86 human immun
4	4028	87.9	1427	2	Q9WF62	Q9wf62 human immun
5	4021	87.7	1429	2	Q6X4P8	Q6x4p8 human immun
6	3993	87.1	1425	2	Q6X4Q6	Q6x4q6 human immun
7	3986	87.0	1427	2	Q9WF96	Q9wf96 human immun
8	3982	86.9	1430	2	Q6X4R4	Q6x4r4 human immun
9	3973.5	86.7	1427	2	Q9WF71	Q9wf71 human immun
10	3969	86.6	1427	2	Q9WFA0	Q9wfa0 human immun
11	3966.5	86.5	1427	2	Q9WF65	Q9wf65 human immun
12	3965	86.5	1437	2	Q9WF77	Q9wf77 human immun
13	3957.5	86.4	1427	2	Q9WF74	Q9wf74 human immun
14	3951.5	86.2	1427	2	Q8ADX5	Q8adx5 human immun
15	3943	86.0	1425	2	Q9WF79	Q9wf79 human immun
16	3936	85.9	1427	2	Q9WF98	Q9wf98 human immun
17	3931.5	85.8	1429	2	Q9WF84	Q9wf84 human immun
18	3922	85.6	1428	2	Q6X4P0	Q6x4p0 human immun
19	3912.5	85.4	1427	2	Q9WF68	Q9wf68 human immun
20	3876.5	84.6	1432	2	Q8ADZ9	Q8adz9 human immun
21	3873	84.5	1433	2	Q8ADE9	Q8ade9 human immun
22	3871	84.5	1434	2	Q8ADV1	Q8adv1 human immun
23	3869	84.4	1430	2	Q7SV36	Q7sv36 human immun
24	3865	84.3	1425	2	Q7SV20	Q7sv20 human immun
25	3859.5	84.2	1433	2	Q8ADN7	Q8adn7 human immun
26	3858.5	84.2	1433	2	Q8ADK5	Q8adk5 human immun
27	3856.5	84.1	1433	2	Q8ADS7	Q8ads7 human immun
28	3853.5	84.1	1433	2	Q8ADG5	Q8adg5 human immun
29	3850.5	84.0	1435	2	Q9DQ33	Q9dq33 human immun
30	3849.5	84.0	1433	2	Q8ADE1	Q8ade1 human immun
31	3844.5	83.9	1431	2	Q8ADR1	Q8adr1 human immun
32	3843.5	83.9	1433	2	Q8ADC5	Q8adc5 human immun
33	3842.5	83.8	1433	2	Q8AE32	Q8ae32 human immun
34	3841.5	83.8	1433	2	Q8ADM1	Q8adm1 human immun
35	3841.5	83.8	1433	2	Q8ADU3	Q8adu3 human immun
36	3840.5	83.8	1437	2	O41782	O41782 human immun
37	3839	83.8	1434	2	Q8Q7Z7	Q8q7z7 human immun
38	3839	83.8	1435	2	Q9WLJ3	Q9wlj3 human immun
39	3838.5	83.8	1433	2	Q8AD85	Q8ad85 human immun
40	3836.5	83.7	1433	2	Q7SPT9	Q7spt9 human immun
41	3832.5	83.6	1433	2	Q8ADW7	Q8adw7 human immun
42	3832.5	83.6	1436	2	Q8AD93	Q8ad93 human immun
43	3830.5	83.6	1433	2	Q8AE23	Q8ae23 human immun
44	3829.5	83.6	1433	2	Q8ADR9	Q8adr9 human immun
45	3827.5	83.5	1433	2	Q8ADH3	Q8adh3 human immun

ALIGNMENTS

```
RESULT 1
Q9WF92
ID   Q9WF92     PRELIMINARY;     PRT;  1427 AA.
AC   Q9WF92;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110975; AAD17135.1; -.
DR   HSSP; P04585; 1JLA.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
```

```
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW   Protease; RNA-directed DNA polymerase; Transferase.
SQ   SEQUENCE   1427 AA;  161059 MW;  F3B0C741A1EFB10A CRC64;

Alignment Scores:
Pred. No.:              2.68e-159      Length:        1427
Score:                  4178.00        Matches:       785
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  99.87%         Mismatches:    0
Query Match:            91.16%         Indels:        0
DB:                     2              Gaps:          0

US-09-610-313B-30 (1-2469) x Q9WF92 (1-1427)

Qy        13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
             :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379

Qy        73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy       133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy       193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439

Qy       253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459

Qy       313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479

Qy       373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy       433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519

Qy       493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy       553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559

Qy       613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy       673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy       733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy       793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy       853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy       913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy       973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy      1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy      1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739

Qy      1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759

Qy      1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy      1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy      1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy      1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy      1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy      1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy     1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899

Qy     1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy     1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy     1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

Qy     1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy     1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy     1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019

Qy     1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy     2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy     2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy     2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy     2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119

Qy     2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1120 GlyIleGlyGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139

Qy     2353 TTCCTGGACGGCATCGAT 2370
            ||||||||||||||||||
Db     1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 2
Q9WF89
ID   Q9WF89      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF89;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110974; AAD17126.1; -.
DR   HSSP; P04585; 1JLA.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  161293 MW;  270F25DF67615C06 CRC64;

Alignment Scores:
Pred. No.:              3.87e-159     Length:         1427
Score:                  4174.00       Matches:        784
Percent Similarity:     99.87%        Conservative:   1
Best Local Similarity:  99.75%        Mismatches:     1
Query Match:            91.08%        Indels:         0
DB:                     2             Gaps:           0

US-09-610-313B-30 (1-2469) x Q9WF89 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
```

```
                 :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759

Qy       1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy       1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899

Qy       1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

Qy       1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019

Qy       1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy       2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy       2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099
```

```
Qy       2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139

Qy       2353 TTCCTGGACGGCATCGAT 2370
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 3
Q9WF86
ID   Q9WF86      PRELIMINARY;      PRT;  1428 AA.
AC   Q9WF86;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110973; AAD17119.1; -.
DR   HSSP; P04585; 1JLA.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1428 AA;  161328 MW;  2D61CD9FEABD2387 CRC64;

Alignment Scores:
Pred. No.:              5.59e-159      Length:        1428
Score:                  4170.00        Matches:       783
Percent Similarity:     100.00%        Conservative:  3
Best Local Similarity:  99.62%         Mismatches:    0
Query Match:            90.99%         Indels:        0
DB:                     2              Gaps:          0

US-09-610-313B-30 (1-2469) x Q9WF86 (1-1428)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetMetGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy      733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy      793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy      853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy      913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy      973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy     1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy     1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739

Qy     1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759

Qy     1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy     1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy     1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy     1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy     1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy     1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy     1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
            ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db      880 LeuGluLeuAlaGluAspArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899

Qy     1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy     1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy     1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

Qy     1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy     1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy     1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
            ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db     1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019

Qy     1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy     2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy     2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy     2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy     2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119

Qy     2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1120 GlyIleGlyGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139

Qy     2353 TTCCTGGACGGCATCGAT 2370
            ||||||||||||||||||
Db     1140 PheLeuAspGlyIleAsp 1145

RESULT 4
Q9WF62
ID   Q9WF62      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF62;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110961; AAD17021.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
```

```
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  160992 MW;  AC2ACBF190C77DAD CRC64;

Alignment Scores:
Pred. No.:                2.63e-153      Length:        1427
Score:                    4028.00        Matches:       750
Percent Similarity:       97.84%         Conservative:  19
Best Local Similarity:    95.42%         Mismatches:    17
Query Match:              87.89%         Indels:        0
DB:                       2              Gaps:          0

US-09-610-313B-30 (1-2469) x Q9WF62 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::|||||||||||   ||||||   |||   |||||:::||||||||||||||||||
Db        360 LeuAlaGluAlaMetGlyGlnAlaAsnSerMetAsnIleMetMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||   |||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db        380 LysAsnProLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              ||||||||||||||||||||||||   ||||||   ||||||||||||||||||||
Db        440 LysAlaArgGluPheProSerGluGlnThrArgAlaIleSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||||||||||||||||    ||||||    |||||
Db        460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluGlyGlnGlyAlaLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
                  ||||||||||||||||||||||||||||||||||:::|||:::|||||||||||||||
Db        480 LeuProGlnIleThrLeuTrpGlnArgProLeuValThrIleArgValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||:::::||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluMetGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 AspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaLeuAsnProGluIleValIle 759

Qy       1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
              ||||||||||||||||:::|||||||||:::|||||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799

Qy       1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819
```

```
Qy       1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        820 ValGlnProIleAsnLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIle 959

Qy       1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
              ||||||||||||||||||||||||||||:::|||||||||:::|||||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluGluGluProIleValGlyAlaGluThrPhe 1019

Qy       1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy       2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ArgGlyArgGlnArgIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy       2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGlyValAsnIleValThrAspSerGln 1079

Qy       2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
              ||||||||||||   ||||||:::||||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyLysGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2353 TTCCTGGACGGCATCGAT 2370
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

RESULT 5
Q6X4P8

```
ID   Q6X4P8      PRELIMINARY;      PRT;  1429 AA.
AC   Q6X4P8;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14709249; DOI=10.1089/088922203771881220;
RA   Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,
RA   McCutchan F.E.;
RT   "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT   in full genome analysis, is not sustained in subgenomic regions.";
RL   AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., McCutchan F.E.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maayan S., Zeira M.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AY255825; AAP76538.1; -.
DR   HSSP; P03367; 1A8G.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
```

```
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   1429 AA;  161586 MW;  D0B9CD124DC64E19 CRC64;

Alignment Scores:
Pred. No.:              5.02e-153     Length:        1429
Score:                  4021.00       Matches:       747
Percent Similarity:     97.97%        Conservative:  27
Best Local Similarity:  94.56%        Mismatches:    12
Query Match:            87.74%        Indels:        4
DB:                     2             Gaps:          1

US-09-610-313B-30 (1-2469) x Q6X4P8 (1-1429)

Qy        13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
             :::||||||||||||||||||   :::||||||||:::||||||:::||||||||
Db       358 LeuAlaGluAlaMetSerGlnAlaAsnAsnAlaAsnIleMetMetGlnLysSerAsnPhe 377

Qy        73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
             ||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       378 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 397

Qy       133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
             ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db       398 AsnCysArgAlaProArgLysArgGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 417

Qy       193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
             |||||||||||||||||||||||||||||||||||:::|||||||||      ||||||
Db       418 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheGlnGlnGly 437

Qy       253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
             :::|||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db       438 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 457

Qy       313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGC--------- 363
             ||||||||||||||||||||    |||||||||||||    ||||||||||||
Db       458 GlnValArgGlyAspAsnProIleSerGluAlaGlyAspGluArgGlnGlyAlaLeuGln 477

Qy       364 ---ACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
                |||||||||    ||||||||||||||||||||||||||||||||||||:::
Db       478 ArgThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIle 497

Qy       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
             ||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db       498 GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGlu 517

Qy       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
             ::::::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       518 IleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 537

Qy       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       538 ValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrVal 557

Qy       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
             |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       558 LeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCys 577

Qy       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       578 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 597

Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
             |||||||||||||||||||||||||||||||||:::|||||||||:::||||||||||||
Db       598 AspGlyProLysValLysGlnTrpProLeuThrLysGluLysIleGluAlaLeuThrAla 617

Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       618 IleCysGluGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyr 637

Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       638 AsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAsp 657

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       658 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 677

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       678 ProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 697

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
             ||||||||||||||||||     |||||||||||||||||||||||||||||||||||||
Db       698 SerValProLeuAspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsn 717

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       718 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 737

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
             |||:::||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       738 ProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnPro 757

Qy      1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
             |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       758 GluLeuValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 777

Qy      1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
             ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db       778 GlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThr 797

Qy      1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       798 ProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro 817

Qy      1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
             ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db       818 AspLysTrpThrValGlnProIleGlnLeuProGluLysGluSerTrpThrValAsnAsp 837

Qy      1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       838 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal 857

Qy      1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       858 ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThr 877

Qy      1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
             ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db       878 GluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGly 897

Qy      1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
             |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db       898 ValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGln 917

Qy      1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       918 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 937

Qy      1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       938 MetArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 957

Qy      1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
             ||||||     ||||||||||||||||||||   ||||||||||||||||||||||||||
Db       958 MetGluGlyIleValIleTrpGlyLysValProLysPheArgLeuProIleGlnLysGlu 977

Qy      1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        978 ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProAspTrpGluPhe 997

Qy       1921 GTGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        998 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGly 1017

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ValGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1037

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 TyrValThrAspArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLys 1057

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1077

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 1097

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
              |||||||||||||||||||||||||||:::||||||:::|||||||||||||||||||
Db       1098 LeuValAsnGlnIleIleGluGlnLeuIleArgLysGluArgValTyrLeuSerTrpVal 1117

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              |||||||||||||||||||||||||||||||||:::||||||||||||||   ||||||
Db       1118 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIle 1137

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              ||||||||||||||||||||||||||||||
Db       1138 ArgLysValLeuPheLeuAspGlyIleAsp 1147

RESULT 6
Q6X4Q6
ID   Q6X4Q6      PRELIMINARY;      PRT;  1425 AA.
AC   Q6X4Q6;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14709249; DOI=10.1089/088922203771881220;
RA   Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,
RA   McCutchan F.E.;
RT   "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT   in full genome analysis, is not sustained in subgenomic regions.";
RL   AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., McCutchan F.E.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maayan S., Zeira M.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AY255824; AAP76529.1; -.
DR   HSSP; Q72498; 1K6Y.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   1425 AA;  161115 MW;  782315D4D5305E18 CRC64;

Alignment Scores:
Pred. No.:              6.59e-152      Length:         1425
Score:                  3993.00        Matches:        738
Percent Similarity:     97.71%         Conservative:   30
Best Local Similarity:  93.89%         Mismatches:     18
Query Match:            87.13%         Indels:         0
DB:                     2              Gaps:           0

US-09-610-313B-30 (1-2469) x Q6X4Q6 (1-1425)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||:::||||||   :::   :::|||:::|||||||||:::||||||
Db        358 LeuAlaGluAlaMetThrGlnAlaAsnAsnThrHisIleMetMetGlnArgAsnAsnPhe 377

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||||||||||||:::||||||||||||||||||||||||||||||:::||||||
Db        378 LysGlyProLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 397

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 417

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
```

```
Db    418 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluIleLeuAlaPheProGlnGly 437

Qy    253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
Db    438 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrAsnGlyGluLeu 457

Qy    313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
Db    458 GlnValArgArgAsnAsnProArgSerLysThrGlyValGluArgGlnGlyThrLeuAsn 477

Qy    373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
Db    478 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 497

Qy    433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
Db    498 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 517

Qy    493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
Db    518 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 537

Qy    553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
Db    538 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 557

Qy    613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
Db    558 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 577

Qy    673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
Db    578 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 597

Qy    733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
Db    598 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 617

Qy    793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db    618 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 637

Qy    853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db    638 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 657

Qy    913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db    658 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 677

Qy    973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db    678 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 697

Qy   1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db    698 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 717

Qy   1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db    718 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 737

Qy   1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db    738 GlnSerSerMetThrArgIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 757

Qy   1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
Db    758 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 777

Qy   1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
Db    778 LysIleGluGluLeuArgGlyHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 797

Qy   1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
Db    798 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 817

Qy   1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
Db    818 ValGlnProIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 837

Qy   1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
Db    838 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 857

Qy   1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
Db    858 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 877

Qy   1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
Db    878 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 897

Qy   1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
Db    898 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyAspAspGlnTrpThrTyrGln 917

Qy   1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
Db    918 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 937

Qy   1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
Db    938 HisThrAsnAspValLysGlnLeuValGlyAlaValGlnLysIleAlaMetGluGlyIle 957

Qy   1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
Db    958 ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 977

Qy   1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
Db    978 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProAspTrpGluPheValAsnThrPro 997

Qy   1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
Db    998 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyValGluThrPhe 1017

Qy   1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
Db   1018 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1037

Qy   2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
Db   1038 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1057

Qy   2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
Db   1058 AlaIleGlnIleAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1077

Qy   2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
Db   1078 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1097

Qy   2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
Db   1098 IleIleGluGlnLeuIleLysLysGluArgIleTyrLeuThrTrpValProAlaHisLys 1117

Qy   2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
Db   1118 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAsnGlyIleArgLysValLeu 1137

Qy   2353 TTCCTGGACGGCATCGAT 2370
Db   1138 PheLeuAspGlyIleAsp 1143
```

```
RESULT 7
Q9WF96
ID   Q9WF96       PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF96;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110976; AAD17144.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  160518 MW;  FBDA73D603A68B18 CRC64;

Alignment Scores:
Pred. No.:              1.26e-151      Length:         1427
Score:                  3986.00        Matches:        734
Percent Similarity:     97.58%         Conservative:   33
Best Local Similarity:  93.38%         Mismatches:     19
Query Match:            86.97%         Indels:         0
DB:                     2              Gaps:           0

US-09-610-313B-30 (1-2469) x Q9WF96 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||   :::   |||||::::::|||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetIleGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::|||   :::||||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||   |||:::||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheGlyGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              :::|||   ||||||||||||||||||   ||||||:::   ||||||||   |||
Db        460 ArgValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||:::
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysValGlyGlyGlnVal 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||::::::||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              |||||||||   |||||||||||||||||||||||||||||||||||||:::||||||
Db        540 AspGlnIleThrIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||:::||||||||||:::||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy      913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy      973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy     1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
            ||||||   ||||||||||||||||||||||||||||||||||||||||     ||||||
Db      700 AspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719

Qy     1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
            ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db      720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy     1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
            |||||||||||||||||||||||||||||||||||:::||||||   |||||||||||||
Db      740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759

Qy     1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy     1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
            ||||||||||||||||||   ||||||||:::||||||||||||||||||||||||||||
Db      780 LysIleGluGluLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799

Qy     1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy     1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
            ||||||||||||:::||||||:::|||:::||||||||||||||||||||||||||||||
Db      820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy     1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy     1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
            ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db      860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879

Qy     1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
            ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db      880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy     1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
            |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db      900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy     1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy     1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
            ||||||||||||||||||||||||   ||||||||||||||||||   ||||||||||||
Db      940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959

Qy     1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy     1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy     1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
            ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db     1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

Qy     1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
            ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db     1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039

Qy     2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
            |||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db     1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy     2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
            ||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db     1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079

Qy     2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy     2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
            ||||||||||||||||||   ||||||::::::|||||||||||||||||||||||||||
Db     1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119

Qy     2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
            |||||||||||||||||||:::||||||||||||||||   |||||||||||||||||||
Db     1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy     2353 TTCCTGGACGGCATCGAT 2370
            ||||||||||||||||||
Db     1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 8
Q6X4R4
ID   Q6X4R4      PRELIMINARY;      PRT;  1430 AA.
AC   Q6X4R4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14709249; DOI=10.1089/088922203771881220;
RA   Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,
RA   McCutchan F.E.;
RT   "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT   in full genome analysis, is not sustained in subgenomic regions.";
RL   AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., McCutchan F.E.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maayan S., Zeira M.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AY255823; AAP76521.1; -.
DR   HSSP; P12497; 1B9D.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
```

```
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   1430 AA;  161540 MW;  5BBA39B85D03FA0B CRC64;

Alignment Scores:
Pred. No.:              1.81e-151     Length:         1430
Score:                  3982.00       Matches:        735
Percent Similarity:     97.59%        Conservative:   36
Best Local Similarity:  93.04%        Mismatches:     15
Query Match:            86.89%        Indels:         4
DB:                     2             Gaps:           1

US-09-610-313B-30 (1-2469) x Q6X4R4 (1-1430)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||   |||:::   ||||||:::||||||:::|||||||||
Db        359 LeuAlaGluAlaMetSerGlnValThrAsnThrAsnIleMetMetGlnLysSerAsnPhe 378

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||::::::   |||||||||||||||||||||||||||||||||:::||||||
Db        379 LysGlyProArgLysThrIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 398

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyArgGluGlyHisGlnMet 418

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||:::|||||||||    ||||||
Db        419 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheGlnGlnGly 438

Qy        253 AAGGCCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||   ||||||   ||||||||||||||||||||||||||||||
Db        439 GluAlaArgGluPheProProGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 458

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC------------CAG 360
              |||||||||||||||||||||   ||||||||||||   ||||||            |||
Db        459 GlnValArgGlyAspAsnProLeuSerGluAlaGlyThrGluArgProGlyThrLeuGln 478

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||   |||||||||||||||||||||||||||||||:::||||||:::
Db        479 GlyThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIle 498

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGlu 518

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ::::::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        519 IleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 538

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||
Db        539 ValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlySerVal 558

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 LeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCys 578

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        579 ThrLeuAsnPheProIleSerProIleGluThrIleProValLysLeuLysProGlyMet 598

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 AspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAla 618

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              |||||||||:::||||||:::||||||||||||:::||||||||||||||||||||||||
Db        619 IleCysGluAspMetGluArgGluGlyLysIleSerLysIleGlyProGluAsnProTyr 638

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        639 AsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAsp 658

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 678

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 ProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 698

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||   |||
Db        699 SerValProLeuAspGluAsnPheArgLysTyrThrAlaPheThrIleProSerThrAsn 718

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 738

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              |||:::|||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        739 ProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnPro 758

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        759 GluIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 778

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              |||||||||||||||:::||||||||||||   ||||||||||||||||||||||||||||
Db        779 GlnHisArgAlaLysValGluGluLeuArgGlyHisLeuLeuArgTrpGlyPheThrThr 798

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 ProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro 818
```

```
Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||     |||:::||||||||||||:::||||||||||||||||||
Db        819 AspLysTrpThrValGlnThrIleGlnLeuProGluLysAspSerTrpThrValAsnAsp 838

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal 858

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        859 ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValAlaLeuThr 878

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        879 GluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGly 898

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        899 ValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGln 918

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        919 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 938

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        939 ArgArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 958

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 MetGluSerIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGlu 978

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        979 ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 998

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        999 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 1018

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1038

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||:::||||||||||||:::|||||||||      |||||||||||||||
Db       1039 TyrValThrAspLysGlyArgGlnLysValValSerLeuAsnGluThrThrAsnGlnLys 1058

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1059 ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1078

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db       1079 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerAspSerGlu 1098

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
              :::|||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db       1099 IleValAsnGlnIleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpVal 1118

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db       1119 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerLysGlyIle 1138

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              ||||||||||||||||||||||||||||||
Db       1139 ArgLysValLeuPheLeuAspGlyIleAsp 1148
```

```
RESULT 9
Q9WF71
ID   Q9WF71      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF71;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110964; AAD17047.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  161326 MW;  B9FC43C29B26CFA2 CRC64;

Alignment Scores:
```

```
Pred. No.:                 3.96e-151     Length:          1427
Score:                     3973.50       Matches:         731
Percent Similarity:        97.46%        Conservative:    35
Best Local Similarity:     93.00%        Mismatches:      19
Query Match:               86.70%        Indels:          1
DB:                        2             Gaps:            1

US-09-610-313B-30 (1-2469) x Q9WF71 (1-1427)

Qy     13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
Db    361 LeuAlaGluAlaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsnPhe 379

Qy     73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
Db    380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy    133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
Db    400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy    193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
Db    420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy    253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
Db    440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy    313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
Db    460 GlnValArgGlyAspAsnProArgSerGluThrArgAlaGluGlyGlnGlyAsnPheAsn 479

Qy    373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
Db    480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy    433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
Db    500 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519

Qy    493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
Db    520 GlyLysTrpLysProLysIleIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy    553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
Db    540 GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy    613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
Db    560 ThrProIleAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy    673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
Db    580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy    733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
Db    600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy    793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db    620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy    853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db    640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy    913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db    660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy    973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db    680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy   1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db    700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719

Qy   1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db    720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy   1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db    740 GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759

Qy   1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
Db    760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy   1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
Db    780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys 799

Qy   1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
Db    800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy   1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
Db    820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy   1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
Db    840 IleGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysHisLeuCys 859

Qy   1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
Db    860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy   1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
Db    880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy   1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
Db    900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy   1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
Db    920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy   1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
Db    940 HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleAlaMetGluCysIle 959

Qy   1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
Db    960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979

Qy   1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
Db    980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy   1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
Db   1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019

Qy   1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
Db   1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039

Qy   2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
```

```
                :::||||||||||||||||||||||   ||||||||||||||||   |||||||||
Db       1040 LysGlyArgGlnLysIleValSerLeuAsnGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099

Qy       2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
              ||||||||||||||||||:::||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2353 TTCCTGGACGGCATCGAT 2370
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 10
Q9WFA0
ID   Q9WFA0      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WFA0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110978; AAD17162.1; -.
DR   PIR; S49086; S49086.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  160503 MW;  24C410797550C0F8 CRC64;

Alignment Scores:
Pred. No.:                6e-151        Length:         1427
Score:                    3969.00       Matches:        731
Percent Similarity:       97.46%        Conservative:   35
Best Local Similarity:    93.00%        Mismatches:     20
Query Match:              86.60%        Indels:         0
DB:                       2             Gaps:           0

US-09-610-313B-30 (1-2469) x Q9WFA0 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||    :::    ||||||:::::::||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetIleGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::|||    :::||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||:::||||||||||||:::||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAspPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||   ||||||||||||||||||||||   ||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerGlyGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||   ||||||||||||||||||||||||      ||||||:::   ||||||||   |||
Db        460 GlnValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnVal 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              |||||||||||||||||||||||||||||||||||||||||||||||:::::::||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539
```

```
Qy      553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
            |||||||||   |||||||||||||||||||||||||||||||||||||||:::||||||
Db      540 AspGlnIleThrIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy      613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
            |||||||||||||||||||||||||||:::|||||||||:::||||||||||||||||||
Db      560 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe 579

Qy      673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy      733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy      793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy      853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy      913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy      973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy     1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
            ||||||   |||||||||||||||||||||||||||||||||||||||   |||||||||
Db      700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719

Qy     1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
            |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db      720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy     1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
            ||||||||||||||||||||||||||||||||||||:::|||   |||||||||||||||
Db      740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759

Qy     1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy     1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
            |||||||||||||||   |||||||||:::||||||||||||||||||||||||||||||
Db      780 LysIleGluGluLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799

Qy     1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy     1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
            ||||||||||||:::||||||:::|||:::||||||||||||||||||||||||||||||
Db      820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy     1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy     1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
            |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db      860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879

Qy     1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
            |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db      880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy     1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
            ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db      900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy     1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy     1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
            |||||||||||||||||||||   ||||||||||||||||||   |||||||||||||||
Db      940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959

Qy     1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
            |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      960 ValLeuTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy     1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
            ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db      980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProLysTrpGluPheValAsnThrPro 999

Qy     1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
            ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db     1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

Qy     1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
            ||||||||||||||||||||||||||||||:::|||||||||||||   ||||||||||||
Db     1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyCysValThrAsp 1039

Qy     2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
            ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db     1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy     2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
            |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db     1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079

Qy     2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy     2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
            |||||||||||||||   ||||||:::|||||||||||||||||||||||||||||||||
Db     1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119

Qy     2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
            |||||||||:::|||:::||||||||||||||||   |||||||||||||||||||||||
Db     1120 GlyIleGlyGlyHisGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy     2353 TTCCTGGACGGCATCGAT 2370
            ||||||||||||||||||
Db     1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 11
Q9WF65
ID   Q9WF65     PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF65;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
```

```
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110962; AAD17030.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW   Protease; RNA-directed DNA polymerase; Transferase.
SQ   SEQUENCE   1427 AA;  161347 MW;  595F19865DE81D63 CRC64;

Alignment Scores:
Pred. No.:              7.55e-151     Length:          1427
Score:                  3966.50       Matches:         732
Percent Similarity:     97.46%        Conservative:    34
Best Local Similarity:  93.13%        Mismatches:      19
Query Match:            86.55%        Indels:          1
DB:                     2             Gaps:            1

US-09-610-313B-30 (1-2469) x Q9WF65 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::|||||||||||||||||   |||:::   |||::::::|||||||||||||||||||
Db        361 LeuAlaGluAlaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::||||||:::|||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||||      ||||||   |||||||||||||||
Db        460 GlnValArgGlyAspAsnProArgSerGluThrArgAlaGluGlyGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              :::||||||||||||||||||||||||||||||||||||||||||::::::::|||||||
Db        500 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              :::||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        540 GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   ||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||   |||||||||||||||||||||   :::||||||:::|||||||||
Db        740 GlnSerSerMetIleLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759

Qy       1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779
```

```
Qy       1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
              ||||||||||||||||:::||||||||||:::||||||   |||||||||   ||||||
Db        780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProTyrLysLys 799

Qy       1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
              ||||||||||:::||||||:::|||:::||||||||||||:::|||||||||||||||
Db        820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValHisAspIleGlnLysLeu 839

Qy       1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
              ||||||||||||||||||||||||||||||||||||||||||:::|||:::   ||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysHisLeuCys 859

Qy       1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
              |||||||||||||:::||||||:::|||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
              ||||||||||||||||||||||||||||||||||   ||||||||||||   ||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleThrMetGluSerIle 959

Qy       1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
              ||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979

Qy       1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019

Qy       1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039

Qy       2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
              :::||||||||||||||||||   ||||||||||||||||||||||:::   |||||||||
Db       1040 LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnArgAlaGluLeuGln 1059

Qy       2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
              |||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099

Qy       2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
              ||||||||||||||||||:::|||||||||||||||   |||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2353 TTCCTGGACGGCATCGAT 2370
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

RESULT 12

```
Q9WF77
ID   Q9WF77      PRELIMINARY;      PRT;  1437 AA.
AC   Q9WF77;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110967; AAD17072.1; -.
DR   HSSP; P24736; 1NCP.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
```

```
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1437 AA;  161978 MW;  65DAA2562FC317A6 CRC64;

Alignment Scores:
Pred. No.:              8.67e-151      Length:         1437
Score:                  3965.00        Matches:        740
Percent Similarity:     96.73%         Conservative:   30
Best Local Similarity:  92.96%         Mismatches:     16
Query Match:            86.52%         Indels:         10
DB:                     2              Gaps:           2

US-09-610-313B-30 (1-2469) x Q9WF77 (1-1437)

Qy      13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
Db     360 LeuAlaGluAlaMetSerGlnAlaAsnSerValAsnIleMetMetGlnLysSerAsnPhe 379

Qy      73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
Db     380 LysGlyProArgArgAsnValLysCysPheAsnCysGlyLysGluGlyHisIleAlaLys 399

Qy     133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
Db     400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy     193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
Db     420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy     253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCC------------------AACAGC 294
Db     440 GluAlaArgGluPheProProGluGlnIleArgAlaSerSerProAsnSerThrAsnSer 459

Qy     295 CCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAG 354
Db     460 ProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGlu 479

Qy     355 ------------CGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
Db     480 GlyGlnGlyThrLeuGlnGlyThrLeuAsnCysProGlnIleThrLeuTrpGlnArgPro 499

Qy     403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
Db     500 LeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAlaAsp 519

Qy     463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
Db     520 AsnThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 539

Qy     523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
Db     540 IleGlyGlyPheIleLysValArgGlnTyrAspGlnIleValIleGluIleCysGlyLys 559

Qy     583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
Db     560 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 579

Qy     643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
Db     580 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 599

Qy     703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
Db     600 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpThrLeuThrGluGluLys 619

Qy     763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
Db     620 IleLysAlaLeuThrGluIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 639

Qy     823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
Db     640 GlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLys 659

Qy     883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
Db     660 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 679

Qy     943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
Db     680 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp 699

Qy    1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
Db     700 MetGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 719

Qy    1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
Db     720 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 739

Qy    1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
Db     740 GlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluPro 759

Qy    1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGC 1242
Db     760 PheArgLeuGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 779

Qy    1243 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1302
Db     780 SerAspLeuGluIleGlyGlnArgArgAlaGlnIleGluGluLeuArgGluHisLeuLeu 799

Qy    1303 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATG 1362
Db     800 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 819

Qy    1363 GGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1422
Db     820 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProAspLysAsp 839

Qy    1423 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1482
Db     840 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 859

Qy    1483 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1542
Db     860 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThr 879

Qy    1543 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1602
Db     880 AspValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu 899

Qy    1603 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1662
Db     900 LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln 919

Qy    1663 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1722
Db     920 LysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 939

Qy    1723 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1782
Db     940 ThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThrGlu 959

Qy    1783 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1842
Db     960 AlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThrProLysPheArg 979

Qy    1843 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1902
Db     980 LeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrp 999
```

```
Qy       1903 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 1019

Qy       1963 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2022
              |||||||||||    |||    ||||||||||||||||||||||||||||||||||||||
Db       1020 LysGluProIleProGlyValGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 1039

Qy       2023 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2082
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 LysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSerLeuThr 1059

Qy       2083 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGly 1079

Qy       2143 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 SerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 1099

Qy       2203 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2262
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db       1100 AspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleGlnLysGluTrp 1119

Qy       2263 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2322
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       1120 ValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLys 1139

Qy       2323 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              |||||||||:::||||||||||||:::||||||||||||||||||||||
Db       1140 LeuValSerGlnGlyIleArgLysMetLeuPheLeuAspGlyIleAsp 1155

RESULT 13
Q9WF74
ID   Q9WF74      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF74;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110965; AAD17056.1; -.
DR   PIR; S49086; S49086.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  161299 MW;  924682CF5095D763 CRC64;

Alignment Scores:
Pred. No.:               1.73e-150     Length:         1427
Score:                   3957.50       Matches:        728
Percent Similarity:      97.33%        Conservative:   37
Best Local Similarity:   92.62%        Mismatches:     20
Query Match:             86.35%        Indels:         1
DB:                      2             Gaps:           1

US-09-610-313B-30 (1-2469) x Q9WF74 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::|||||||||||||||||||   :::   :::::::::||||||:::|||||||||
Db        361 LeuAlaGluAlaMetSerGlnAla---AsnThrSerValMetMetGlnLysSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::||||||:::|||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||    |||||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||         |||     ||||||       |||
Db        460 GlnValArgGlyAspAsnProArgSerGluThrArgValGluGlyGlnGlyAsnPheAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy          433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
                :::||||||||||||||||||||||||||||||||||||||||||:::::::::||||||
Db          500 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519

Qy          493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy          553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
                :::||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          540 GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy          613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
                ||||||||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db          560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnPheGlyCysThrLeuAsnPhe 579

Qy          673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy          733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db          600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp 619

Qy          793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy          853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy          913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy          973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy         1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
                ||||||   ||||||||||||||||||||||||||||||||||||||||||   ||||||
Db          700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719

Qy         1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
                ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db          720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy         1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
                ||||||||||||||||||||||||||||||||||||   :::||||||:::|||||||||
Db          740 GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759

Qy         1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy         1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
                ||||||||||||||||||:::||||||||||:::||||||   ||||||||||||||||||
Db          780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys 799

Qy         1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy         1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
                ||||||||||||:::||||||:::|||:::||||||||||||||||||||||||||||||
Db          820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy         1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
                ||||||||||||||||||||||||||||||||||||||||||:::||||||   ||||||
Db          840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValArgHisLeuCys 859

Qy         1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy         1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
                ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db          880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy         1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
                |||||||||||||||:::||||||:::|||||||||||||||||||||||||||||||||
Db          900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy         1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy         1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
                |||||||||||||||||||||||||||||||   |||:::||||||||||||||||||||
Db          940 HisThrAsnAspValLysGlnLeuThrGluValValArgLysIleAlaMetGluSerIle 959

Qy         1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
                ||||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db          960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979

Qy         1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy         1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
                ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db         1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019

Qy         1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
                ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||:::
Db         1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 1039

Qy         2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
                :::|||||||||||||||   ||||||||||||||||||||||   |||||||||
Db         1040 LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy         2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy         2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db         1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099

Qy         2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
                |||||||||||||||||||||||||||::::::|||||||||||||||||||||||||||
Db         1100 IleIleGluGlnLeuIleLysLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119

Qy         2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
                |||||||||||||||||:::|||||||||||||||   ||||||||||||||||||||||
Db         1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy         2353 TTCCTGGACGGCATCGAT 2370
                ||||||||||||||||||
Db         1140 PheLeuAspGlyIleAsp 1145

RESULT 14
Q8ADX5
ID   Q8ADX5      PRELIMINARY;      PRT;  1427 AA.
AC   Q8ADX5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
GN   Name=gag-pol;
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22375625; PubMed=12487816; DOI=10.1089/088922202320886325;
RA   Harris M.E., Serwadda D., Sewankambo N., Wabwire F., Kim B.,
RA   Kigozi G., Kiwanuka N., Phillips J.B., Meehen M., Lutalo T.,
RA   Lane J.R., Merling R., Gray R., Wawer M., Birx D.L., Robb M.L.,
RA   McCutchan F.E.;
RT   "Among 46 near full length HIV type 1 genome sequences from Rakai
RT   District, Uganda, subtype D and AD recombinants predominate.";
RL   AIDS Res. Hum. Retroviruses 18:1281-1290(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L., Robb M.L.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., Phillips J.B., Lane J.R., Merling R., McCutchan F.E.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Lutalo T.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Meehen M., Wawer M.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Serwadda S., Sewankambo N., Wabwire F., Kigozi G., Kiwanuka N.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF484491; AAN73556.1; -.
DR   HSSP; P24736; 1NCP.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW   Protease; RNA-directed DNA polymerase; Transferase.
FT   NON_TER       1      1
SQ   SEQUENCE   1427 AA;  161058 MW;  20559041E2A88B5C CRC64;

Alignment Scores:
Pred. No.:              3e-150        Length:         1427
Score:                  3951.50       Matches:        728
Percent Similarity:     97.20%        Conservative:   36
Best Local Similarity:  92.62%        Mismatches:     21
Query Match:            86.22%        Indels:         1
DB:                     2             Gaps:           1

US-09-610-313B-30 (1-2469) x Q8ADX5 (1-1427)

Qy          13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
               :::||||||||||||||||||            :::||||||:::||||||:::      |||
Db         361 LeuAlaGluAlaMetSerGlnValGlnHisSerAsnIleMetMetGlnGlnGlyLysPhe 380

Qy          73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
               :::|||   ||||||   |||||||||||||||||||||||||||||||||:::||||||
Db         381 ArgGlyGlnLysArg---IleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 399

Qy         133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy         193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
               ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db         420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy         253 AAGGCCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
               :::|||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db         440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy         313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
               ||||||||||||||||||||||||   |||||||||||||||   |||||||||||||||
Db         460 GlnValArgGlyAspAsnProArgPheGluAlaGlyAlaGluGlyGlnGlyThrLeuAsn 479

Qy         373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
               |||||||||||||||||||||||||||   |||:::|||:::||||||||||||:::
Db         480 PheProGlnIleThrLeuTrpGlnArgProPheValAlaIleArgValGlyGlyGlnLeu 499

Qy         433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
               ||||||||||||||||||||||||||||||||||||||||||||||||:::   ||||||
Db         500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleLysLeuPro 519

Qy         493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
               |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         520 GlyAsnTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy         553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
               ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db         540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy         613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy         673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy         733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db         600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp 619
```

```
Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIle 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysArgLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              |||:::||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        740 GlnAsnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 759

Qy       1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
              ||||||||||||||||||||||||||||||||||||:::|||      ||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuAspIleGluGlnHisArgAla 779

Qy       1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy       1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
              |||||||||:::||||||||||||:::|||||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys 859

Qy       1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
              |||||||||||||||:::|||:::||||||||||||   |||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGln 919

Qy       1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysValArgThrAla 939

Qy       1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
              ||||||||||||:::|||||||||||||||||||||||||||||||||   |||||||||
Db        940 HisThrAsnAspIleLysGlnLeuThrGluAlaValGlnLysIleAlaArgGluSerIle 959

Qy       1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 IleIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluAla 979

Qy       1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
              ||||||||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleProGlyValGluThrPhe 1019

Qy       1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysMetGlyLysAlaGlyTyrValThrAsp 1039

Qy       2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
              |||||||||||||||:::|||||||||||||||||||||||||||   ||||||||||||
Db       1040 ArgGlyArgGlnLysIleIleSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnValValThrAspSerGln 1079

Qy       2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnLys 1099

Qy       2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
              ||||||||||||||||||:::|||||||||||||||   |||||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2353 TTCCTGGACGGCATCGAT 2370
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 15
Q9WF79
ID   Q9WF79      PRELIMINARY;      PRT;  1425 AA.
AC   Q9WF79;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110969; AAD17087.1; -.
DR   HSSP; P24736; 1NCP.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
```

```
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1425 AA;  160569 MW;  6DC8A1AEF3C88DA7 CRC64;

Alignment Scores:
Pred. No.:              6.55e-150    Length:         1425
Score:                  3943.00      Matches:        731
Percent Similarity:     97.33%       Conservative:   34
Best Local Similarity:  93.00%       Mismatches:     19
Query Match:            86.04%       Indels:         2
DB:                     2            Gaps:           1

US-09-610-313B-30 (1-2469) x Q9WF79 (1-1425)

Qy        13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
             :::||||||||||||||||||      :::::||||||:::||||||::::::||||||
Db       360 LeuAlaGluAlaMetSerGlnAlaAsnAsnSerAsnIleMetMetGlnLysAsnAsnPhe 379

Qy        73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
             |||||||||   |||||:::||||||||||||||||||||||||||||||||||||||||
Db       380 LysGlyProThrArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy       133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
             ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db       400 AsnCysLysAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy       193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
             ||||||||||||||||||:::|||||||||||||||||||:::|||||||||||||||||
Db       420 LysAspCysThrGluArgArgAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy       253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
             :::||||||||||||||||||||||||      |||||||||||||||||||:::|||
Db       440 GluAlaArgGluPheProSerGluGlnThrGlyAlaAsnSerProThrSerArgLysLeu 459

Qy       313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
                   ||||||:::||||||   ||||||||||||   ||||||||||||||||||
Db       460 ------ArgGlyAsnAsnProCysSerGluAlaGlyAspGluArgGlnGlyThrLeuAsn 477

Qy       373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       478 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 497

Qy       433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
             ||||||||||||||||||||||||||||||||||||||||||||||||::::::||||||
Db       498 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 517

Qy       493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
             |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db       518 GlyLysTrpArgProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 537

Qy       553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
             |||||||||   ||||||||||||||||||||||||||||||||||||:::||||||
Db       538 AspGlnIleProIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 557

Qy       613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
             |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db       558 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPhe 577

Qy       673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       578 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 597

Qy       733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       598 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 617

Qy       793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       618 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 637

Qy       853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       638 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 657

Qy       913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       658 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 677

Qy       973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       678 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 697

Qy      1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
             ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       698 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 717

Qy      1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       718 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 737

Qy      1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
             |||||||||||||||||||||||||||||||||||||||:::||||||||||:::||||||
Db       738 GlnSerSerMetThrLysIleLeuGluProPheArgAlaLysAsnProGluLeuValIle 757

Qy      1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       758 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 777

Qy      1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
             :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       778 ArgIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 797

Qy      1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       798 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 817

Qy      1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
             ||||||   |||:::|||||||||||||||||||||||||||||||||||||||||||||
Db       818 ValGlnSerIleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 837

Qy      1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
```

```
                 |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        838 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyValLysValArgGlnLeuCys 857

Qy       1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        858 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProProThrGluGluAlaGlu 877

Qy       1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        878 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 897

Qy       1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
              |||||||||||||||:::|||||||||||||||||||||   ||||||||||||||||||
Db        898 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlyAspGlnTrpThrTyrGln 917

Qy       1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 937

Qy       1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        938 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleSerMetGluSerIle 957

Qy       1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
              ||||||||||||      ||||||||||||||||||||||||||||||   ||||||   
Db        958 ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluAlaTrpGluAla 977

Qy       1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 997

Qy       1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
              ||||||||||||||||||||||||||||||   ||||||:::   |||||||||||||||
Db        998 ProLeuValLysLeuTrpTyrGlnLeuGluThrGluProMetAlaGlyAlaGluThrPhe 1017

Qy       1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1037

Qy       2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
              :::|||||||||::::::|||:::||||||||||||||||||||||||   |||||||||
Db       1038 LysGlyArgGlnGluValValThrLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1057

Qy       2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
              ||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db       1058 AlaIleGlnLeuAlaLeuGlnAspSerGlyProGluValAsnIleValThrAspSerGln 1077

Qy       2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1097

Qy       2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1117

Qy       2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
              |||||||||||||||::::::||||||||||||||||||   |||||||||:::||||||
Db       1118 GlyIleGlyGlyAsnGluLysValAspLysLeuValSerSerGlyIleArgGluValLeu 1137

Qy       2353 TTCCTGGACGGCATCGAT 2370
              ||||||||||||||||||
Db       1138 PheLeuAspGlyIleAsp 1143
```

Search completed: June 2, 2005, 05:02:46
Job time : 301.175 secs

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  2, 2005, 02:53:39 ; Search time 40.9328 Seconds
                                          (without alignments)
                                          11607.276 Million cell updates/sec

Title:          US-09-610-313B-30
Perfect score:  4583
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2469

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:      566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155137_15159/app_query.fasta_1.7893
-DB=PIR_79 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_189_@runat_31052005_155137_15159 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      PIR_79:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3543.5	77.3	1002	2	S54378	pol polyprotein -
2	3535.5	77.1	1002	1	GNLJND	HIV-1 retropepsin
3	3530	77.0	1003	1	GNVWLV	HIV-1 retropepsin
4	3513	76.7	1015	1	GNVWH3	HIV-1 retropepsin
5	3491	76.2	1003	2	T09440	pol polyprotein -
6	3488	76.1	1003	1	B44001	HIV-1 retropepsin
7	3487	76.1	1003	1	GNVWA2	HIV-1 retropepsin
8	3477	75.9	1012	1	GNVWVL	HIV-1 retropepsin
9	3249	70.9	1027	1	GNLJSI	HIV-1 retropepsin
10	3138	68.5	912	2	S33980	pol polyprotein -
11	3132	68.3	902	2	T01668	pol polyprotein -
12	2809	61.3	559	2	B47175	reverse transcript
13	2789	60.9	559	2	A47175	reverse transcript
14	2388	52.1	1039	2	S46347	pol polyprotein -
15	2244	49.0	1055	2	S53092	pol polyprotein -
16	2240.5	48.9	1055	1	GNLJST	HIV-1 retropepsin
17	2231.5	48.7	1056	1	GNLJG3	HIV-1 retropepsin
18	2231	48.7	1036	1	GNLJG2	HIV-1 retropepsin
19	2223	48.5	1034	1	GNLJCA	HIV-1 retropepsin
20	2222.5	48.5	1035	1	GNLJGG	HIV-1 retropepsin
21	2212.5	48.3	1019	2	T11560	pol polyprotein -
22	2206.5	48.1	1009	2	S28081	pol polyprotein -
23	2185.5	47.7	1032	2	S12153	pol polyprotein -
24	2108	46.0	1058	2	S08436	pol polyprotein -
25	2088	45.6	1061	1	GNLJG4	HIV-1 retropepsin
26	2054.5	44.8	1054	1	GNLJG5	HIV-1 retropepsin
27	1998.5	43.6	656	2	S30484	pol polyprotein -
28	1969.5	43.0	656	2	S30483	pol polyprotein -
29	1428	31.2	1124	2	S23820	pol polyprotein -
30	1425	31.1	1124	1	GNLJFP	HIV-1 retropepsin
31	1382	30.2	1124	2	B45557	HIV-1 retropepsin
32	1172	25.6	1146	1	GNLJ22	HIV-1 retropepsin
33	1172	25.6	1146	1	GNLJEW	HIV-1 retropepsin
34	1169.5	25.5	1109	1	B45345	HIV-1 retropepsin
35	1152.5	25.1	1087	2	JQ1162	Pol protein - Maed
36	1152.5	25.1	1145	1	GNLJEV	HIV-1 retropepsin
37	1149.5	25.1	1086	1	B46335	HIV-1 retropepsin
38	1142	24.9	1101	1	B45390	HIV-1 retropepsin
39	1138	24.8	1101	1	GNLJVS	HIV-1 retropepsin
40	1128	24.6	216	2	S43127	reverse transcript
41	1109	24.2	219	2	S32152	RNA-directed DNA p
42	1105	24.1	219	2	S32118	RNA-directed DNA p
43	1105	24.1	219	2	S32139	RNA-directed DNA p
44	1103	24.1	219	2	S32089	RNA-directed DNA p
45	1103	24.1	219	2	S32160	RNA-directed DNA p

ALIGNMENTS

```
RESULT 1
S54378
pol polyprotein - human immunodeficiency virus type 1 (fragment)
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C;Accession: S54378
R;Theodore, T.; Buckler-White, A.J.
submitted to the EMBL Data Library, July 1989
A;Reference number: S54377
A;Accession: S54378
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1-1002 <THE>
A;Cross-references: UNIPROT:P12499; EMBL:M22639; NID:g329377; PIDN:AAA45366.1; PID:g3293
C;Superfamily: pol polyprotein.
C;Keywords: polyprotein

Alignment Scores:
Pred. No.:              7.65e-156       Length:         1002
Score:                  3543.50         Matches:        659
Percent Similarity:     96.25%          Conservative:   34
Best Local Similarity:  91.53%          Mismatches:     24
Query Match:            77.32%          Indels:         3
DB:                     2               Gaps:           2

US-09-610-313B-30 (1-2469) x S54378 (1-1002)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||||||||||||||      |||      ||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
                 |||||||||||||||||||||||||||:::|||    |||   ||||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTC------CCCCAGATCACCCTG 390
              |||||   ||||||||||||||||||||:::::|||        ||||||||||||||
Db         41 SerGluThrGlyAlaGluArgGlnGlyThrValSerPheAsnCysProGlnIleThrLeu 60
```

```
Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              |||||||||||||||:::||||||:::|||||||||:::|||||||||||||||||||||
Db         61 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||   |||||||||||||||||||||:::||||||||||||||||||||||||
Db        121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              |||||||||||||||||||||||||||      ||||||   ||||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||:::::::::|||||||||||||||||||||||||||:::||||||||||||||||||
Db        201 IleSerArgValGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230
              ||||||||||||||||||   :::||||||||||||||||||||||||||||||||||||
Db        321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy       1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290
              |||||||||||||||||||||||||||||||||||||||      |||||||||||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy       1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy       1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1410
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||:::|||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400

Qy       1411 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy       1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr 440

Qy       1531 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1590
              |||||||||:::::::::||||||||||||||||||||||||||||||||||||||||||
Db        441 LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460

Qy       1591 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1650
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1651 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1710
              |||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy       1711 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1770
              |||||||||||||||||||||||||||:::||||||   |||||||||||||||||||||
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520

Qy       1771 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1830
              ||||||   |||   ||||||||||||:::   |||||||||||||||||||||||||||
Db        521 GlnLeuAlaGluValValGlnLysIleSerThrGluSerIleValIleTrpGlyLysThr 540

Qy       1831 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890
              ||||||||||||||||||||||||||||||||||||||||||||||||   :::||||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpValGluTyrTrp 560

Qy       1891 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1951 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       2011 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2070
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2071 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
              |||      |||:::||||||||||||||||||||||||||||||   ||||||||||||
Db        621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640

Qy       2131 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2190
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2191 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2250
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680

Qy       2251 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2310
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              |||:::|||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 2
GNLJND
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate NDK)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA polym
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: JQ0067
R;Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.;
Gene 81, 275-284, 1989
A;Title: Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunoc
A;Reference number: JQ0065; MUID:90034200; PMID:2806917
A;Accession: JQ0067

```
A;Molecule type: DNA
A;Residues: 1-1002 <SPI>
A;Cross-references: UNIPROT:P18802; GB:M27323; NID:g328154; PIDN:AAA44869.1; PID:g328158
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;56-154/Product: retropepsin #status predicted <RTP>
F;80/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:
Pred. No.:              1.79e-155      Length:          1002
Score:                  3535.50        Matches:         659
Percent Similarity:     96.25%         Conservative:    34
Best Local Similarity:  91.53%         Mismatches:      24
Query Match:            77.14%         Indels:          3
DB:                     1              Gaps:            2

US-09-610-313B-30 (1-2469) x GNLJND (1-1002)

Qy       220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
             |||||||||||||||||||||||||||||||||||||||||    ||||||   |||||||||
Db         1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaGlyGluPheSerSerGluGln 20

Qy       280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
                ||||||||||||||||||||||||||||:::|||         |||||||||||
Db        21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspAsnProLeu 40

Qy       337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTG 390
             ||||||   |||||||||||||||||||:::       :::||||||||||||||||||
Db        41 SerGluThrGlyAlaGluArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60

Qy       391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
             |||||||||||||||:::||||||:::||||||||||:::|||||||||||||||||||
Db        61 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy       451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
             ||||||||||||||||||||||||||||||:::::||||||||||||||||||||||||
Db        81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy       511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy       571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
             |||||||||   ||||||:::||||||||||||||||||||||||||||||||||||||
Db       121 IleCysGlyTyrLysAlaMetGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy       631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
             |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db       141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy       691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy       751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
             |||||||||||||||||||||||||||     ||||||     ||||||||||||||||||
Db       181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200

Qy       811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
             |||::::::|||||||||||||||||||||||||||||:::||||||||||||||||||
Db       201 IleSerArgIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220

Qy       871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy       931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy       991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy      1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy      1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
             ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db       301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy      1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230
             ||||||||||||||||||    :::||||||||||||||||||||||||||||||||||
Db       321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy      1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290
             |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db       341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy      1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350
             :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy      1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1410
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleAsnLeu 400

Qy      1411 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1470
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy      1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530
             ||||||||||||||||    ||||||||||||:::|||||||||||||||||||||||
Db       421 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 440

Qy      1531 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1590
             |||||||||:::::||||||||||||||||||||||||||||||||||||||||||||||
Db       441 LysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460

Qy      1591 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1650
             ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||:::
Db       461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy      1651 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1710
             ||||||:::|||||||||||       |||||||||||||||||||||||||||||||
Db       481 AlaGluLeuGlnLysGlnGlyAspGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy      1711 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1770
             ||||||||||||||||||||||||:::   |||    |||||||||||||||||||||
Db       501 LysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspValLys 520

Qy      1771 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1830
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       521 GlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 540

Qy      1831 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890
             ||||||||:::|||||||||||||||||||||||||||||||||||   :::||||||
Db       541 ProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpIleGluTyrTrp 560

Qy      1891 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1950
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy      1951 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2010
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy      2011 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2070
             ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||:::
Db       601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy      2071 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
```

```
Db        621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640

Qy       2131 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2190

Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2191 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2250

Db        661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680

Qy       2251 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2310

Db        681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370

Db        701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 3
GNVWLV
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LAV-1a)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 03-Jun-2002
C;Accession: A03966
R;Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.
Cell 40, 9-17, 1985
A;Title: Nucleotide sequence of the AIDS virus, LAV.
A;Reference number: A90866; MUID:85099333; PMID:2981635
A;Accession: A03966
A;Molecule type: DNA
A;Residues: 1-1003 <WAI>
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;57-155/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status experimental

Alignment Scores:

Pred. No.:	3.2e-155	Length:	1003
Score:	3530.00	Matches:	659
Percent Similarity:	95.98%	Conservative:	33
Best Local Similarity:	91.40%	Mismatches:	25
Query Match:	77.02%	Indels:	4
DB:	1	Gaps:	2

US-09-610-313B-30 (1-2469) x GNVWLV (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279

Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333

Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387

Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447

Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507

Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567

Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627

Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687

Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747

Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807

Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867

Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927

Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987

Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047

Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107

Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167

Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227

Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287

Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347

Db        361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407

Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467

Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527

Db        421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587

Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
```

```
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707

Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767

Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827

Db        521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887

Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947

Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007

Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067

Db        601 AlaSerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127

Db        621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187

Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247

Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307

Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367

Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370

Db        721 Asp 721
```

RESULT 4
GNVWH3
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate HTLV-III
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C;Accession: A03965
R;Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K.J.; Starcich, B.; Josephs, S.F.; Dora
nberger, J.A.; Papas, T.S.; Ghrayeb, J.; Chang, N.T.; Gallo, R.C.; Wong-Staal, F.
Nature 313, 277-284, 1985
A;Title: Complete nucleotide sequence of the AIDS virus, HTLV-III.
A;Reference number: A93353; MUID:85111123; PMID:2578615
A;Accession: A03965
A;Molecule type: DNA
A;Residues: 1-1015 <RAT>
A;Cross-references: UNIPROT:P03366; GB:M15654; GB:K02008; GB:K02009; GB:K02010; NID:g326
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;69-167/Product: retropepsin #status predicted <RTP>
F;93/Active site: Asp (shared with dimeric partner) #status experimental

```
Alignment Scores:
Pred. No.:              1.94e-154     Length:        1015
Score:                  3513.00       Matches:       658
Percent Similarity:     94.27%        Conservative:  33
Best Local Similarity:  89.77%        Mismatches:    26
Query Match:            76.65%        Indels:        16
DB:                     1             Gaps:          3
```

US-09-610-313B-30 (1-2469) x GNVWH3 (1-1015)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267

Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303

Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357

Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411

Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471

Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531

Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591

Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651

Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711

Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771

Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831

Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891

Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951

Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011

Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
```

```
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTG 1251
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1252 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1311
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1312 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1371
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1431
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1491
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1492 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1551
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1552 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1611
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1612 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1671
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1672 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1732 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1791
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1851
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1852 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1911
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1912 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1972 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2031
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2032 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2091
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2092 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2151
Db        641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2211
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2212 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2271
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2272 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2332 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 5
T09440
pol polyprotein - human immunodeficiency virus type 1 (strain JRFL) (fragment)
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T09440
R;Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie,
submitted to the EMBL Data Library, July 1996
A;Reference number: Z16673
A;Accession: T09440
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1003 <PAN>
A;Cross-references: UNIPROT:Q75755; EMBL:U63632; NID:g1465777; PID:g1465779
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein

```
Alignment Scores:
Pred. No.:              1.99e-153     Length:        1003
Score:                  3491.00       Matches:       650
Percent Similarity:     95.56%        Conservative:  39
Best Local Similarity:  90.15%        Mismatches:    28
Query Match:            76.17%        Indels:        4
DB:                     2             Gaps:          2

US-09-610-313B-30 (1-2469) x T09440 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db         21 ThrArgAlaAsnSerProThrArgLysGluLeuGlnValTrpGlyArgAspSerAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACC 387
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyAsnValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db         81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
```

```
Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||||        ||||||   ||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysArgLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
              ||||||||||||||||||      :::||||||:::|||:::||||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeu 360

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
              ||||||||||||||||      |||||||||||:::||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
                 ||||||||||:::::::|||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
              :::|||||:::||||||||||||          |||||||||||||||||||||||||||
Db        481 IleAlaGluLeuGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
              ||||||    ||||||||||||||||||||:::   |||   ||||||||||||||||||
Db        501 PheLysIleLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
              |||||||||||||||||||||||||||||||||      ||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
                 |||||||||:::||||||||||||||||||||||||||||||||||||:::|||
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
              ||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
              ||||||||||||||||:::||||||||||||||||||||:::||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2068 ATCGTGAGCCTGACCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
              :::|||||||||:::|||||||||||||||||||||||||||||||||     ||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
              |||||||||||||||     |||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
              |||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
              ||||||:::||||||||||||||     ||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370
              |||
Db        721 Asp 721
```

RESULT 6
B44001
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (strain YU-2)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: B44001
R;Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.
J. Virol. 66, 6587-6600, 1992
A;Title: Complete nucleotide sequence, genome organization, and biological properties of
A;Reference number: A44001; MUID:93021387; PMID:1404605
A;Accession: B44001
A;Molecule type: DNA
A;Residues: 1-1003 <LIY>
A;Cross-references: UNIPROT:P35963; GB:M93258
C;Comment: This protein is synthesized as a gag-pol polyprotein.
C;Comment: Specific enzymatic cleavages may yield mature proteins including proteinase,
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;57-155/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status predicted

```
Alignment Scores:
Pred. No.:              2.74e-153    Length:         1003
Score:                  3488.00      Matches:        651
Percent Similarity:     95.42%       Conservative:   37
Best Local Similarity:  90.29%       Mismatches:     29
Query Match:            76.11%       Indels:         4
DB:                     1            Gaps:           2

US-09-610-313B-30 (1-2469) x B44001 (1-1003)

Qy     220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db       1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy     280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db      21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40

Qy     334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db      41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy     388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db      61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy     448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db      81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy     508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db     101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy     568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db     121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy     628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db     141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy     688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db     161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy     748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db     181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy     808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db     201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy     868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db     221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy     928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db     241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy     988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db     261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuHisGluAspPheArg 280

Qy    1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db     281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyThrArgTyrGln 300

Qy    1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db     301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy    1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
Db     321 ThrIleLeuGluProPheArgLysGlnAsnProAspLeuValIleTyrGlnTyrMetAsp 340

Qy    1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
Db     341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy    1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
Db     361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy    1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
Db     381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy    1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
Db     401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy    1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
Db     421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy    1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
Db     441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy    1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
Db     461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy    1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
Db     481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy    1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
Db     501 PheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520

Qy    1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
Db     521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy    1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
Db     541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy    1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
Db     561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy    1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
Db     581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600

Qy    2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
Db     601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnLysGlyArgGlnLys 620

Qy    2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
Db     621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy    2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
Db     641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy    2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
Db     661 IleGlnAlaGlnProAspArgSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
```

```
Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
              ||||||:::|||||||||||||||   |||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370
              |||
Db        721 Asp 721
```

RESULT 7
GNVWA2
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate ARV-2)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C;Accession: A03968
R;Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Sh
Science 227, 484-492, 1985
A;Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).
A;Reference number: A04003; MUID:85090453; PMID:2578227
A;Accession: A03968
A;Molecule type: DNA
A;Residues: 1-1003 <SAN>
A;Cross-references: UNIPROT:P03369; GB:K02007; NID:g328658; PIDN:AAB59876.1; PID:g328662
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle
F;57-147/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status predicted

```
Alignment Scores:
Pred. No.:              3.04e-153      Length:          1003
Score:                  3487.00        Matches:         650
Percent Similarity:     95.42%         Conservative:    38
Best Local Similarity:  90.15%         Mismatches:      29
Query Match:            76.09%         Indels:          4
DB:                     1              Gaps:            2

US-09-610-313B-30 (1-2469) x GNVWA2 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||   ||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------GACAACCCC 333
                 |||||||||||||||||   |||||||||||||||   |||         :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||||:::||||||||||||:::      |||||||||||||||||
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              ||||||||||||||||||:::|||:::||||||||||||:::||||||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   :::
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProVal 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              ||||||||||||   |||||||||||||||||||||:::|||||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              |||||||||||||||||||||||||||      ||||||   ||||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              |||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
              ||||||||||||||||||   :::||||||:::|||||||||||||||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
              ||||||||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMet 400

Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
              ||||||||||||||||||   ||||||||||||:::||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
                |||||||||||:::::::::||||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
              ||||||||||||||:::||||||||||||||   ||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
              |||||||||||||||||||||||||      |||||||||||||||||||||||||||||
Db        481 ValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500
```

```
Qy      1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
Db       501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy      1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
Db       521 LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys 540

Qy      1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
Db       541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyr 560

Qy      1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
Db       561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy      1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
Db       581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy      2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
Db       601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy      2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
Db       621 ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy      2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
Db       641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy      2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
Db       661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy      2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
Db       681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy      2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
Db       701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720

Qy      2368 GAT 2370
Db       721 Asp 721
```

RESULT 8
GNVWVL
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LV)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C;Accession: A03967
R;Muesing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.
Nature 313, 450-458, 1985
A;Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi
A;Reference number: A93355; MUID:85111157; PMID:2982104
A;Accession: A03967
A;Molecule type: DNA
A;Residues: 1-1012 <MUE>
A;Cross-references: UNIPROT:P03368
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;66-164/Product: retropepsin #status predicted <RTP>
F;90/Active site: Asp (shared with dimeric partner) #status experimental

Alignment Scores:

Pred. No.:	8.79e-153	Length:	1012
Score:	3477.00	Matches:	651
Percent Similarity:	94.22%	Conservative:	34
Best Local Similarity:	89.55%	Mismatches:	26
Query Match:	75.87%	Indels:	16
DB:	1	Gaps:	3

US-09-610-313B-30 (1-2469) x GNVWVL (1-1012)

```
Qy       238 GCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------------------------ 267
Db         4 AlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerPro 23

Qy       268 ------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG--- 318
Db        24 ThrIleSerSerGluGlnThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrp 43

Qy       319 ---CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------ 369
Db        44 GlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPhe 63

Qy       370 AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAG 429
Db        64 AsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGln 83

Qy       430 ATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTG 489
Db        84 LeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeu 103

Qy       490 CCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAG 549
Db       104 ProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGln 123

Qy       550 TACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGC 609
Db       124 TyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIleGlyThrValLeuValGly 143

Qy       610 CCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAAC 669
Db       144 ProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsn 163

Qy       670 TTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCC 729
Db       164 PheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyPro 183

Qy       730 AAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAG 789
Db       184 LysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThr 203

Qy       790 GAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCC 849
Db       204 GluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrPro 223

Qy       850 GTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAG 909
Db       224 ValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGlu 243

Qy       910 CTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGC 969
Db       244 LeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGly 263

Qy       970 CTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCC 1029
Db       264 LeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValPro 283

Qy      1030 CTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACC 1089
Db       284 LeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThr 303

Qy      1090 CCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATC 1149
Db       304 ProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIle 323

Qy      1150 TTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTG 1209
```

```
                ||||||||||||||||||||||||||||||||||||||    :::||||||:::||||||
Db        324 PheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleVal 343

Qy       1210 ATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGC 1269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 IleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArg 363

Qy       1270 GCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAG 1329
                 ||||||||||||||||:::|||||||||||||||||   ||||||||||||||
Db        364 ThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLys 383

Qy       1330 AAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGG 1389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 LysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrp 403

Qy       1390 ACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAG 1449
              ||||||||||||||   ||||||||||||:::||||||||||||||||||||||||||||
Db        404 ThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLys 423

Qy       1450 CTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTG 1509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 LeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeu 443

Qy       1510 TGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCC 1569
              ||||||||||||||||||   ||||||||||||:::::::::||||||||||||||||||
Db        444 CysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAla 463

Qy       1570 GAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTAC 1629
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        464 GluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyr 483

Qy       1630 GACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTAC 1689
              ||||||||||||||||:::|||||||||||||||||||||    |||||||||||
Db        484 AspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyr 503

Qy       1690 CAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACC 1749
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        504 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 523

Qy       1750 GCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGC 1809
              |||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        524 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSer 543

Qy       1810 ATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAG 1869
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        544 IleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGluThrTrpGlu 563

Qy       1870 ACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACC 1929
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        564 ThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThr 583

Qy       1930 CCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACC 1989
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        584 ProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThr 603

Qy       1990 TTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACC 2049
              |||||||||||||||||||||||||||||||:::::::::||||||||||||||||||||
Db        604 PheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGlyLysAlaGlyTyrValThr 623

Qy       2050 GACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTG 2109
              ::::::||||||||||||:::|||   ||||||   ||||||||||||||||||||||||
Db        624 AsnLysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeu 643

Qy       2110 CAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGC 2169
              |||||||||   |||||||||||||||||||||   ||||||||||||||||||||||||
Db        644 GlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSer 663

Qy       2170 CAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAAC 2229
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        664 GlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsn 683

Qy       2230 CAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCAC 2289
              |||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        684 GlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHis 703

Qy       2290 AAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2349
              ||||||||||||||||||||||:::|||||||||||||||   ||||||||||||:::
Db        704 LysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysIle 723

Qy       2350 CTGTTCCTGGACGGCATCGAT 2370
              |||||||||||||||||||||
Db        724 LeuPheLeuAspGlyIleAsp 730
```

RESULT 9
GNLJSI
HIV-1 retropepsin (EC 3.4.23.16) - simian immunodeficiency virus SIVcpz
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: simian immunodeficiency virus SIVcpz
A;Note: host Pan troglodytes (chimpanzee)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: S09984
R;Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.
Nature 345, 356-359, 1990
A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
A;Reference number: S09983; MUID:90259077; PMID:2188136
A;Accession: S09984
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1027 <HUE>
A;Cross-references: UNIPROT:P17283; EMBL:X52154
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;81-180/Product: retropepsin #status predicted <RTP>
F;105/Active site: Asp (shared with dimeric partner) #status predicted

```
Alignment Scores:
Pred. No.:               2.73e-142     Length:         1027
Score:                   3249.00       Matches:        599
Percent Similarity:      89.46%        Conservative:   63
Best Local Similarity:   80.95%        Mismatches:     75
Query Match:             70.89%        Indels:         3
DB:                      1             Gaps:           1

US-09-610-313B-30 (1-2469) x GNLJSI (1-1027)

Qy        159 CTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAA 218
              ||||||   |||            |||   |||:::|||   |||||||||     |||:::
Db          7 LeuLeuAlaValTrpAlaArgGlyThrProAsnGluArgLeuHisArgLysThrGlyGlu 26

Qy        219 CTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCA 278
                |||||||||||   ||||||||||||||   :::||||||:::        :::|||||
Db         27 -PhePheArgGluArgLeuAlaPheProGlnArgGluAlaArgGlnLeuCysAlaGluGl 46

Qy        279 GAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------GACAACCC 332
              ||||||   |||   ||||||   |||||||||   |||   |||        :::   ||
Db         46 nAsnArgThrAsnGlyProThrAspArgGluLeuTrpValProGlyGlyArgGluGluPr 66

Qy        333 CCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTG 392
              |         |||   |||   |||:::         :::   |||   |||||||||||||||
Db         66 oGlyGluGluArgGlyArgGluGlnSerIleSerThrAsnLeuProGlnIleThrLeuTr 86

Qy        393 GCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACAC 452
              ||||||||||||:::   :::||||||   ||||||:::   ||||||||||||||||
Db         86 pGlnArgProLeuIleProValLysValGluGlyGlnLeuCysGluAlaLeuLeuAspTh 106

Qy        453 CGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGAT 512
              ||||||||||||||||||:::|||      :::   |||   |||   ||||||||||||||
Db        106 rGlyAlaAspAspThrValIleGluArgIleGlnLeuGlnGlyLeuTrpLysProLysMe 126

Qy        513 GATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGAT 572
              |||||||||||||||||||||||||||||:::|||:::|||      :::      |||||||
Db        126 tIleGlyGlyIleGlyGlyPheIleLysValLysGlnPheAspAsnValHisIleGluIl 146
```

```
Qy     573 CTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGG 632
Db     146 eGluGlyArgLysValValGlyThrValLeuValGlyProThrProValAsnIleIleGl 166

Qy     633 CCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGAC 692
Db     166 yArgAsnIleLeuThrGlnLeuGlyCysThrLeuValPheProIleSerSerIleGluTh 186

Qy     693 CGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGAC 752
Db     186 rValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuSe 206

Qy     753 CGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGAT 812
Db     206 rAlaGluLysIleLysAlaLeuThrGluIleCysGlnGluMetGluLysGluGlyLysIl 226

Qy     813 CACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGA 872
Db     226 eSerLysIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAs 246

Qy     873 CAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTT 932
Db     246 pSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPh 266

Qy     933 CTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGAC 992
Db     266 eTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValTh 286

Qy     993 CGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTA 1052
Db     286 rValLeuAspValGlyAspAlaTyrPheSerCysProLeuAspLysAspPheArgLysTy 306

Qy    1053 CACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAA 1112
Db     306 rThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyValArgTyrGlnTyrAs 326

Qy    1113 CGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGAT 1172
Db     326 nValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSerMetThrLysIl 346

Qy    1173 CCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCT 1232
Db     346 eLeuGluProPheArgGluLysAsnProAspIleThrIleTyrGlnTyrMetAspAspLe 366

Qy    1233 GTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAA 1292
Db     366 uTyrValGlySerAspLeuGluIleAspGlnHisArgLysLysValGluGluLeuArgGl 386

Qy    1293 GCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTT 1352
Db     386 nHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPh 406

Qy    1353 CCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCC 1412
Db     406 eLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuPr 426

Qy    1413 CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGC 1472
Db     426 oGluLysGluValTrpThrValAsnAspIleGlnLysLeuIleGlyLysLeuAsnTrpAl 446

Qy    1473 CAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAA 1532
Db     446 aSerGlnIleTyrProGlyIleLysIleLysGlnLeuCysLysLeuIleArgGlyThrLy 466

Qy    1533 GGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCG 1592
Db     466 sLysLeuThrAspValValProLeuThrProGluAlaGluLeuGluLeuAlaGluAsnAr 486

Qy    1593 CGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGC 1652
Db     486 gGluIleValSerThrProValHisGlyValTyrTyrAspProAspLysGluLeuIleAl 506

Qy    1653 CGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAA 1712
Db     506 aGluIleGlnLysGlnGlyAsnCysGlnTrpThrTyrGlnIlePheGlnGluProHisLy 526

Qy    1713 GAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCA 1772
Db     526 sAsnLeuLysThrGlyLysTyrAlaArgGlnArgSerAlaHisThrAsnAspIleArgGl 546

Qy    1773 GCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCC 1832
Db     546 nLeuAlaGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThrPr 566

Qy    1833 CAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCA 1892
Db     566 oLysPheArgLeuProValGlnLysGluSerTrpGluAlaTrpTrpAlaGluTyrTrpGl 586

Qy    1893 GGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTA 1952
Db     586 nAlaThrTrpIleProGluTrpGluPheIleAsnThrProProLeuValLysLeuTrpTy 606

Qy    1953 CCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAA 2012
Db     606 rSerLeuGluThrGluProIleProThrThrAspThrTyrTyrValAspGlyAlaAlaAs 626

Qy    2013 CCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGT 2072
Db     626 nArgGluThrLysThrGlyLysAlaGlyTyrValThrAspLysGlyLysGlnLysIleIl 646

Qy    2073 GAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCA 2132
Db     646 eSerLeuGluAsnThrThrAsnGlnGlnAlaGluLeuLysAlaLeuLeuLeuAlaLeuGl 666

Qy    2133 GGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCA 2192
Db     666 nAspSerAspGlnGlnValAsnIleValThrAspSerGlnTyrValLeuGlyIleIleGl 686

Qy    2193 GGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAA 2252
Db     686 nSerGlnProAspHisSerGluSerGluLeuValAsnGlnIleIleGluGluLeuIleLy 706

Qy    2253 GAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCA 2312
Db     706 sLysGluLysIleTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGl 726

Qy    2313 GATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db     726 nValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 745
```

RESULT 10
S33980
pol polyprotein - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C;Accession: S33980
R;Carlini, F.
submitted to the EMBL Data Library, November 1991
A;Reference number: S33979
A;Accession: S33980
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-912 <CAR>
A;Cross-references: UNIPROT:Q78241; EMBL:Z11530; NID:g60192; PIDN:CAA77622.1; PID:g60194
C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	3.43e-137	Length:	912
Score:	3138.00	Matches:	578
Percent Similarity:	96.51%	Conservative:	30
Best Local Similarity:	91.75%	Mismatches:	22
Query Match:	68.47%	Indels:	0
DB:	2	Gaps:	0

US-09-610-313B-30 (1-2469) x S33980 (1-912)

```
Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db          1 MetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 20

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db         21 ValLysGlnTyrAspGlnIleAlaIleGluIleCysGlyHisLysAlaIleGlyThrVal 40

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db         41 LeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCys 60

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db         61 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 80

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db         81 AspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuIleGlu 100

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db        101 IleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyr 120

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db        121 AsnThrProValPheAlaIleLysLysLysAspGlyThrLysTrpArgLysLeuValAsp 140

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db        141 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 160

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db        161 ProGlyGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 180

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db        181 SerValProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsn 200

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db        201 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 220

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db        221 ProAlaIlePheGlnAlaSerMetThrLysIleLeuGluProPheArgLysGlnAsnPro 240

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Db        241 GluIleIleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 260

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Db        261 GlnHisArgThrLysIleGluGluLeuArgArgHisLeuLeuArgTrpGlyPheThrThr 280

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Db        281 ProAspLysLysHisGlnLysGluProProPheLeuTrpIleGlyTyrGluLeuHisPro 300

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Db        301 AspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAsp 320

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Db        321 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLysVal 340

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Db        341 ArgGlnLeuCysArgLeuLeuArgGlyAlaLysAlaLeuThrGluValIleProLeuThr 360

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Db        361 LysGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysThrProValHisGly 380

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Db        381 ValTyrTyrAspProSerLysAspLeuValAlaGluIleGlnLysGlnGlyLeuGlyGln 400

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db        401 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 420

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Db        421 MetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 440

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Db        441 ThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGlu 460

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Db        461 ThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 480

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Db        481 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 500

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Db        501 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly 520

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Db        521 TyrValThrAspArgGlyArgGlnLysValIleSerLeuThrAspThrThrAsnGlnLys 540

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Db        541 ThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleVal 560

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Db        561 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 580

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Db        581 LeuValSerGlnIleIleGluHisLeuIleLysLysGluLysValTyrLeuAlaTrpVal 600

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Db        601 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIle 620

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db        621 ArgLysValLeuPheLeuAspGlyIleAsp 630
```

RESULT 11
T01668
pol polyprotein - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T01668
R;Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.
Cell 46, 63-74, 1986
A;Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol
A;Reference number: Z14389; MUID:86245056; PMID:2424612
A;Accession: T01668
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-902 <ALI>
A;Cross-references: EMBL:K03456; NID:g60228; PIDN:CAA28012.1; PID:g60230
C;Superfamily: pol polyprotein

Alignment Scores:

```
Pred. No.:              6.48e-137    Length:          902
Score:                  3132.00      Matches:         578
Percent Similarity:     97.74%       Conservative:    28
Best Local Similarity:  93.23%       Mismatches:      14
Query Match:            68.34%       Indels:          0
DB:                     2            Gaps:            0

US-09-610-313B-30 (1-2469) x T01668 (1-902)

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 20

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||||||||||||||||||||:::|||:::||||||||||||||||||||||||
Db         21 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 40

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db         41 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 60

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db         61 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 80

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              |||||||||||||||||||||||||||   ||||||::::::||||||||||||||||||
Db         81 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 100

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 120

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        121 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 140

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 160

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 180

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 200

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        201 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 220

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230
              ||||||||||||||||||   :::||||||||||||||||||||||||||||||||||||
Db        221 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 240

Qy       1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        241 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 260

Qy       1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350
              :::|||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 280

Qy       1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        281 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 300

Qy       1411 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1470
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 320

Qy       1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        321 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 340

Qy       1531 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1590
              ||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db        341 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 360

Qy       1591 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1650
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||:::
Db        361 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 380

Qy       1651 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1710
              |||||||||||||||||||||      |||||||||||||||||||||||||||   :::
Db        381 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 400

Qy       1711 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1770
              |||||||||||||||||||||||||||::::::::::::|||||||||||||||||||||
Db        401 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 420

Qy       1771 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1830
              ||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db        421 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 440

Qy       1831 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890
              |||||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db        441 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 460

Qy       1891 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 480

Qy       1951 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2010
              ||||||||||||   |||||||||:::|||||||||||||||||||||||||||||||||
Db        481 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 500

Qy       2011 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2070
              |||||||||||||||   |||||||||||||||||||||||||||||||||||||||:::
Db        501 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 520

Qy       2071 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        521 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 540

Qy       2131 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 560

Qy       2191 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2250
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        561 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 580

Qy       2251 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2310
              :::|||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 600

Qy       2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              |||:::|||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        601 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 620

RESULT 12
B47175
reverse transcriptase, AZT-sensitive variant - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: B47175
R;Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993
A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the
A;Reference number: A47175; MUID:93126353; PMID:7678340
A;Accession: B47175
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: DNA
```

```
A;Residues: 1-559 <MOH>
A;Cross-references: UNIPROT:Q9PXX1
A;Note: sequence extracted from NCBI backbone (NCBIP:122099)
C;Superfamily: pol polyprotein

Alignment Scores:
Pred. No.:               4.39e-122      Length:         559
Score:                   2809.00        Matches:        518
Percent Similarity:      96.78%         Conservative:   23
Best Local Similarity:   92.67%         Mismatches:     18
Query Match:             61.29%         Indels:         0
DB:                      2              Gaps:           0

US-09-610-313B-30 (1-2469) x B47175 (1-559)

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 20

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||      ||||||   |||
Db         21 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGlu 40

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db         41 MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal 60

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 80

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 100

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 120

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 140

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        141 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 160

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||    :::||||||:::||||||||
Db        161 GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleValIle 180

Qy       1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr 200

Qy       1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
              ||||||||||||||||:::|||||||||||||||||||||||      ||||||||||||
Db        201 LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLys 220

Qy       1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 240

Qy       1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
              ||||||||||||      ||||||||||||:::|||||||||||||||||||||||||||
Db        241 ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 260

Qy       1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 280

Qy       1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
              |||||||||||||||      ||||||||||||:::::::::||||||||||||||||||
Db        281 LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGlu 300

Qy       1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        301 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 320

Qy       1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
              ||||||||||||:::||||||||||||||||||||||||        |||||||||||||
Db        321 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGln 340

Qy       1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||   |||
Db        341 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla 360

Qy       1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
              |||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        361 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle 380

Qy       1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
              ||||||||||||||||||:::|||:::|||||||||||||||||||||||||||||||||
Db        381 ValIleTrpGlyLysThrProArgPheLysLeuProIleGlnLysGluThrTrpGluThr 400

Qy       1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
              |||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 TrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 420

Qy       1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        421 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 440

Qy       1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||:::
Db        441 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 460

Qy       2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
              :::||||||||||||:::|||     |||||    |||||||||||||||||||||||||
Db        461 LysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeuGln 480

Qy       2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
              ||||||   ||||||||||||||||||||    |||||||||||||||||||||||||||
Db        481 AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln 500

Qy       2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        501 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln 520

Qy       2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
              |||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        521 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys 540

Qy       2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2349
              ||||||||||||||||||||:::|||||||||||||||    |||||||||||||||
Db        541 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 559

RESULT 13
A47175
reverse transcriptase, AZT-resistant variant - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A47175
R;Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993
A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the
A;Reference number: A47175; MUID:93126353; PMID:7678340
A;Accession: A47175
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: nucleic acid
A;Residues: 1-559 <MOH>
A;Cross-references: UNIPROT:Q9PXX2
A;Note: sequence extracted from NCBI backbone (NCBIP:122100)
C;Superfamily: pol polyprotein

Alignment Scores:
```

```
Pred. No.:                 3.66e-121     Length:         559
Score:                     2789.00       Matches:        514
Percent Similarity:        96.60%        Conservative:   26
Best Local Similarity:     91.95%        Mismatches:     19
Query Match:               60.86%        Indels:         0
DB:                        2             Gaps:           0

US-09-610-313B-30 (1-2469) x A47175 (1-559)

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
Db          1 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 20

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
Db         21 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGlu 40

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db         41 MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal 60

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db         61 PheAlaIleLysLysLysAspSerThrArgTrpArgLysLeuValAspPheArgGluLeu 80

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db         81 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 100

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db        101 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 120

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db        121 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 140

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db        141 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 160

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db        161 GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspMetValIle 180

Qy       1213 TACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1272
Db        181 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr 200

Qy       1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1332
Db        201 LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 220

Qy       1333 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1392
Db        221 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 240

Qy       1393 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452
Db        241 ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 260

Qy       1453 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
Db        261 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys 280

Qy       1513 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1572
Db        281 LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleGlnLeuThrGluGluAlaGlu 300

Qy       1573 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1632
Db        301 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 320

Qy       1633 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1692
Db        321 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGln 340

Qy       1693 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
Db        341 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla 360

Qy       1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1812
Db        361 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle 380

Qy       1813 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1872
Db        381 ValIleTrpGlyLysIleProArgPheLysLeuProIleGlnLysGluThrTrpGluAla 400

Qy       1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1932
Db        401 TrpTrpIleGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 420

Qy       1933 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1992
Db        421 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 440

Qy       1993 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2052
Db        441 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 460

Qy       2053 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2112
Db        461 LysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGln 480

Qy       2113 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2172
Db        481 AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln 500

Qy       2173 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2232
Db        501 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln 520

Qy       2233 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2292
Db        521 IleIleGluGluLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys 540

Qy       2293 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2349
Db        541 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 559

RESULT 14
S46347
pol polyprotein - simian immunodeficiency virus SIVagm (isolate SAB-1)
C;Species: simian immunodeficiency virus SIVagm
A;Variety: isolate SAB-1
C;Date: 25-Dec-1994 #sequence_revision 14-Feb-1997 #text_change 26-Aug-1999
C;Accession: S46347
R;Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.; Barre-Sinoussi, F.; Hirsch, V.M.; A
EMBO J. 13, 2935-2947, 1994
A;Title: Mosaic genome structure of simian immunodeficiency virus from West African greer
A;Reference number: S46335; MUID:94298785; PMID:8026477
A;Accession: S46347
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1039 &lt;JIN&gt;
A;Cross-references: EMBL:U04005; NID:g466229; PIDN:AAA21505.1; PID:g466231
A;Experimental source: isolate SAB-1; sabaeus monkey
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
A;Note: this reading frame extends between two stop codons and does not begin with a sta
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein

Alignment Scores:
Pred. No.:              1.13e-102      Length:        1039
```

```
Score:                    2388.00       Matches:        444
Percent Similarity:       74.60%        Conservative:   120
Best Local Similarity:    58.73%        Mismatches:     150
Query Match:              52.11%        Indels:         42
DB:                       2             Gaps:           6

US-09-610-313B-30 (1-2469) x S46347 (1-1039)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC------AAGGCCCGCGAGTTCCCCAGC 273
              |||||||||                  :::|||   |||      :::   :::||||||||||||
Db          1 PhePheArg---------ValTrpProLeuGlyGlnArgGluThrGlnGluPheProSer 17

Qy        274 GAGCAGAACCGCGCCAAC---AGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------ 324
              :::  ::::::  |||  ||||||           ||||||      |||
Db         18 AspLeuHisGlnThrAsnSerSerProAsnGlyThrGlyLeuGlnGlnAlaGlyGlyLys 37

Qy        325 ------------------GACAACCCCGCAGCGAGGCC--------------------- 345
                                :::   |||:::   |||
Db         38 LeuValCysArgGlnThrSerAspGlnArgThrArgAlaArgArgSerSerAsnSerPro 57

Qy        346 ---------------------------GGCGCCGAGCGC--------------------- 357
                                         ||||||
Db         58 ValLysAlaValCysCysSerGlyGluThrAlaGluThrAlaValAlaLysProLeuAla 77

Qy        358 ---------------CAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
                             :::|||   |||      |||||::::::||||||:::||||||
Db         78 ThrThrGluProLeuArgGlyGlyLeuGlnLeuProGlnValSerLeuTrpArgArgPro 97

Qy        403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
              :::   ::::::   :::   ||||||         |||||||||||||||||||||
Db         98 MetLysThrValTyrIleGluGlyGlnLysValThrAlaLeuLeuAspThrGlyAlaAsp 117

Qy        463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
              |||:::|||::::::   :::   |||         |||||||||::::::||||||||
Db        118 AspSerValIleGlnGlyIleGluLeuGlyAspAsnTrpLysProArgIleIleGlyGly 137

Qy        523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
              |||||||||   |||   |||:::   |||               ::::::|||   |||
Db        138 IleGlyGlyCysIleAsnValLysAlaTyrHisAsnGlnGluValLysIleGluAspLys 157

Qy        583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
                       |||:::|||:::|||   |||||||||||||||||||||||||:::
Db        158 ThrCysLysAlaThrIleLeuValGlyGluThrProValAsnIleIleGlyArgAsnVal 177

Qy        643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
              |||   |||||||||   |||||||||               ||||||   :::   |||
Db        178 LeuAlaGlnLeuGlyValThrLeuAsnLeuThrGlnArgGluIleGluProIleLysVal 197

Qy        703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
              ||||||||||||   |||||||||:::::::::||||||||||||::::::||||||
Db        198 HisLeuLysProGlyGlnAspGlyProArgIleArgGlnTrpProLeuSerLysGluLys 217

Qy        763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
              |||:::||||||   ||||||||||||:::::::||||||:::|||   :::   :::|||
Db        218 IleGluAlaLeuLysAlaIleCysGluAspLeuGluLysGlnGlyHisLeuGluArgIle 237

Qy        823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
              |||||||||||||||||||||||||||||||||||||:::|||||||||   |||:::
Db        238 GlyProGluAsnProTyrAsnThrProValPheAlaIleArgLysLysAspLysThrGln 257

Qy        883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
              ||||||   |||:::|||||||||:::|||||||||   ||||||||||||   ||||||
Db        258 TrpArgIleLeuMetAspPheArgGlnLeuAsnLysSerThrGlnAspPheGlnGluVal 277

Qy        943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
              ||||||||||||||||||||||||||||||:::::::::::   :::||||||||||||
Db        278 GlnLeuGlyIleProHisProAlaGlyLeuGlnGlnArgGluGlnIleThrValLeuAsp 297

Qy       1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
              :::||||||||||||||||||   |||||||||   ||||||:::||||||||||||||||
Db        298 IleGlyAspAlaTyrPheSerCysProLeuAspProAspPheGlnLysTyrThrAlaPhe 317

Qy       1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
              ||||||||||||:::||||||   ||||||||||||||||||||   |||||||||
Db        318 ThrIleProSerValAsnAsnArgGluProGlyIleArgTyrGlnTyrLysValLeuPro 337

Qy       1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
              ||||||||||||||||||||:::||||||||:::::::      |||||||||:::
Db        338 GlnGlyTrpLysGlySerProThrIlePheGlnThrThrAlaAsnLysIleLeuGlnGlu 357

Qy       1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGC 1242
              ||||||   :::||||||::::::   |||||||||||||||||||||:::   :::
Db        358 PheArgGlnLysAsnProAspValAspIleTyrGlnTyrMetAspAspMetLeuIleAla 377

Qy       1243 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1302
              ||||||            :::|||         :::::::::||||||   :::|||
Db        378 SerAspArgProLysAlaGluHisLeuValMetValGlnGlnLeuArgAspTyrLeuGlu 397

Qy       1303 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATG 1362
                 |||||||||   ||||||:::||||||   ||||||:::||||||:::||||||||
Db        398 ThrTrpGlyPheLysThrProGluLysLysPheGlnLysAspProProTyrLeuTrpMet 417

Qy       1363 GGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1422
              ||||||||||||:::|||   ||||||      :::|||   |||   |||||||||:::||||||
Db        418 GlyTyrGluLeuTyrProLysLysTrpGlnLeuGlnGluIleThrLeuProGluArgGlu 437

Qy       1423 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1482
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 GluTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 457

Qy       1483 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1542
              |||   |||||||||   :::   |||||||:::|||:::||||||||||:::   ||||||
Db        458 TyrThrGlyIleLysThrLysHisLeuCysArgLeuIleArgGlyAlaArgProLeuThr 477

Qy       1543 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1602
              :::||||||      ||||||||||||||||||||||||   |||||||||:::||||||
Db        478 GluIleValGlnTrpThrGluGluAlaGluLeuGluLeuGluGluAsnArgGlnIleLeu 497

Qy       1603 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1662
              |||:::      |||   ||||||||||||:::      |||   |||::::::
Db        498 ArgGlnLysGlnGlnGlyGlnTyrTyrAspProAlaLeuProLeuArgAlaLysValLeu 517

Qy       1663 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1722
              |||   |||      ||||||   |||||||||||||||      |||   |||
Db        518 LysLeuGlyAspGlyGlnTrpGlyTyrGlnIleTyrGlnProGluAsnLysIleLeuLys 537

Qy       1723 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1782
                 |||||||||||||||::::::||||||||||||||||:::::::::   |||
Db        538 ValGlyLysTyrAlaLysIleLysThrAlaHisThrAsnGluLeuArgMetLeuAlaGly 557

Qy       1783 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1842
                 ||||||||||||   |||||||||||||||||||||:::   |||
Db        558 LeuValGlnLysIleGlyLysGluSerIleValIleTrpGlyGlnIleProIleMetGlu 577

Qy       1843 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1902
              ||||||:::::::::|||   ||||||   ||||||:::||||||||||||   ||||||
Db        578 LeuProValGluArgGluLeuTrpGluGlnTrpTrpSerAspTyrTrpGlnValThrTrp 597

Qy       1903 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1962
              |||||||||||||||   ||||||:::||||||   ||||||:::||||||||||||:::|||
Db        598 IleProGluTrpGluMetValSerThrProGlnLeuIleArgLeuTrpTyrLysLeuVal 617

Qy       1963 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2022
              |||:::||||||   |||         :::||||||||||||||||||||||||   :::
Db        618 LysAspProIleProGlyGluAlaValTyrTyrValAspGlyAlaAlaAsnArgAsnSer 637

Qy       2023 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2082
              |||   ||||||||||||||:::||||||||||||   ||||||:::||||||:::|||
Db        638 LysGluGlyLysAlaGlyTyrLeuThrAspArgGlyAspGlnLysValValAlaLeuGlu 657

Qy       2083 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2142
                 |||||||||||||||   ||||||:::||||||   ||||||||||:::||||||||||
Db        658 AsnThrThrAsnGlnLysAlaGluLeuGluAlaIleLeuLeuAlaLeuArgAspSerGly 677

Qy       2143 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2202
              |||:::|||||||||:::|||||||||||||||||:::|||||||||         :::|||
```

```
Db        678 SerLysValAsnIleIleThrAspSerGlnTyrAlaMetGlyIleIleAlaGlyGluPro 697

Qy       2203 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2262
Db        698 ThrGluSerAspAsnAsnIleValGlnGlnIleIleGluGluLeuIleLysLysGluAla 717

Qy       2263 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2322
Db        718 ValTyrIleAlaTrpValProAlaHisLysGlyValGlyGlyAsnGluGluIleAspLys 737

Qy       2323 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db        738 LeuValSerGlnGlyIleArgGlnValLeuPheLeuAspArgIleGlu 753
```

RESULT 15
S53092
pol polyprotein - human immunodeficiency virus type 2
C;Species: human immunodeficiency virus type 2, HIV-2
C;Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C;Accession: S53092
R;Becker, M.; Zorr, B.; Becker, A.; Habermehl, K.O.
submitted to the EMBL Data Library, March 1995
A;Description: Molecular and phylogenetic characterisation of a Guinea-Bissau-derived hu
A;Reference number: S53091
A;Accession: S53092
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1055 <BEC>
A;Cross-references: UNIPROT:O73194; EMBL:Z48731
C;Superfamily: pol polyprotein
C;Keywords: polyprotein

Alignment Scores:

Pred. No.:	4.8e-96	Length:	1055
Score:	2244.00	Matches:	421
Percent Similarity:	71.11%	Conservative:	128
Best Local Similarity:	54.53%	Mismatches:	184
Query Match:	48.96%	Indels:	39
DB:	2	Gaps:	9

US-09-610-313B-30 (1-2469) x S53092 (1-1055)

```
Qy        156 GGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCAC---CGAGCGCCA 212
Db          3 GlyLeuLeuGluMetTrpGln------------AspArgThrTyrHisGlyLysValPro 18

Qy        213 GGCCAACTT------------------------------------------CTTCCG 227
Db         19 ArgGlnThrGlyGlyPhePheArgAspTrpProLeuGlyLysGluAlaProGlnLeuPro 38

Qy        228 CGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAA------ 281
Db         39 ArgGlyProGlySerAlaGlyAlaAsnThrAsnSerThrProSerArgSerSerSerGly 58

Qy        282 CCGCGCCAACAGCCCCACCAGCCGCGA------GCTGCAGGTGCGCGGCGACAACCCCCG 335
Db         59 ProThrGlyGluIleTyrAlaAlaArgGluLysAlaGluGlyAlaGluArgGluThrIle 78

Qy        336 CAGCGAGGCCGGCGC------CGAGCGCCA------GGGCACCCTGAACTT--------- 374
Db         79 GlnArgGlyAspArgGlyLeuAlaAlaProArgAlaGlyLysAspThrMetGlnGlyAsp 98

Qy        375 ------------------CCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db         99 AsnArgGlyPheAlaAlaPro-GlnPheSerLeuTrpAsnArgProValValThrAlaHi 118

Qy        417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db        118 sIleGluGlyGlnProValGluValLeuLeuAspThrGlyAlaAspAspSerIleValAl 138

Qy        477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db        138 aGlyIleGluLeuGlySerAsnTyrSerProLysIleValGlyGlyIleGlyGlyPheIl 158

Qy        537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db        158 eAsnThrLysGluTyrLysAsnValGluIleGluValLeuGlyLysArgValArgAlaTh 178

Qy        597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db        178 rIleMetThrGlyAspThrProIleAsnIlePheGlyArgAsnIleLeuThrAlaLeuGl 198

Qy        657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db        198 yMetSerLeuAsnLeuProValAlaLysIleGluProIleLysIleMetLeuLysProGl 218

Qy        717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db        218 yLysAspGlyProLysLeuArgGlnTrpProLeuThrLysGluLysIleGluAlaLeuLy 238

Qy        777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db        238 sGluIleCysGluLysMetGluArgGluGlyGlnLeuGluGluAlaProProThrAsnPr 258

Qy        837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db        258 oTyrAsnThrProThrPheAlaIleArgLysLysAspLysAsnLysTrpArgMetLeuIl 278

Qy        897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db        278 eAspPheArgGluLeuAsnLysValThrGlnAspPheThrGluIleGlnLeuGlyIlePr 298

Qy        957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db        298 oHisProAlaGlyLeuAlaLysLysArgArgIleThrValLeuAspValGlyAspAlaTy 318

Qy       1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db        318 rPheSerIleProLeuHisGluAspPheArgGlnTyrThrAlaPheThrLeuProSerVa 338

Qy       1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db        338 lAsnAsnAlaGluProGlyLysArgTyrIleTyrLysValLeuProGlnGlyTrpLysGl 358

Qy       1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db        358 ySerProAlaIlePheGlnTyrThrMetArgGlnValLeuGluProPheArgLysAlaAs 378

Qy       1197 CCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGAT 1256
Db        378 nGlnAspValIleIleIleGlnTyrMetAspAspIleLeuIleAlaSerAspArgThrAs 398

Qy       1257 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1316
Db        398 pLeuGluHisAspArgValValLeuGlnLeuLysGluLeuLeuAsnSerLeuGlyPheSe 418

Qy       1317 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCA 1376
Db        418 rThrProAspGluLysPheGlnLysAspProProTyrArgTrpMetGlyTyrGluLeuTr 438

Qy       1377 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1436
Db        438 pProThrLysTrpLysLeuGlnLysIleGlnLeuProGlnLysGluValTrpThrValAs 458

Qy       1437 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1496
Db        458 nAspIleGlnLysLeuValGlyValLeuAsnTrpAlaAlaGlnIleTyrProGlyIleLy 478

Qy       1497 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1556
Db        478 sThrLysHisLeuCysArgLeuIleArgGlyLysMetThrLeuThrGluGluIleGlnTr 498

Qy       1557 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1616
Db        498 pThrGluLeuAlaGluAlaGluLeuGluGluAsnArgValIleLeuSerGlnGluGlnGl 518
```

```
Qy      1617 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1676
              |||  ||||||         |||:::|||  |||  :::||||||          ::
Db       518 uGlyHisTyrTyrGlnGluGluLysGluLeuGluAlaThrValGlnLysAspGlnAspAs 538

Qy      1677 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1736
             :|||||||||||:::::::::|||         |||  ||||||  ||||||||||
Db       538 nGlnTrpThrTyrLysValHisGlnGlyGlu---LysThrLeuLysValGlyLysTyrAl 557

Qy      1737 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1796
             ||||::::::      ||||||||  |||:::  |||   :::   |||||||||
Db       557 aLysValLysAsnThrHisThrAsnGlyValArgLeuLeuAlaGlnValValGlnLysIl 577

Qy      1797 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1856
             |      |||::::::|||||||||||:::   |||||||||  ||||||::::::::
Db       577 eGlyLysGluAlaLeuValIleTrpGlyArgIleProLysPheHisLeuProValGluAr 597

Qy      1857 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1916
             :|||  ||||||  ||||||  |||||||||||   ||||||||||||:::||||::
Db       597 gGluIleTrpGluGlnTrpTrpAspAspTyrTrpGlnValThrTrpIleProAspTrpAs 617

Qy      1917 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1976
             :||||||:::|||||||||||||||:::|||   :::   |||   |||:::||||||
Db       617 pPheValSerThrProProLeuValArgLeuValPheAsnLeuValLysAspProIlePr 637

Qy      1977 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2036
               ||||||||||||||||   ||||||:::   ||||||::::::|||   ||||||||
Db       637 oGlyAlaGluThrPheTyrThrAspGlySerCysAsnArgGlnSerLysGluGlyLysAl 657

Qy      2037 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2096
             ||||||||||||||||:::||||||   |||:::       |||   :::|||||||||||
Db       657 aGlyTyrValThrAspLysGlyArgAspLysValLysValLeuGluGlnThrThrAsnGl 677

Qy      2097 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2156
             |:::   ||||||:::|||        :::||||||   |||||||||   :::   |||||
Db       677 nGlnAlaGluLeuGluAlaPheAlaMetAlaLeuThrAspSerGlyProLysAlaAsnIl 697

Qy      2157 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2216
             |:::   ||||||||||||      :::||||||:::      ||||||   :::||||||::
Db       697 eIleValAspSerGlnTyrValMetGlyIleValValGlyGlnProThrGluSerGluAs 717

Qy      2217 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2276
             :   :::||||||||||||||||||::::::||||||||||||   :::|||::::::||
Db       717 nArgIleValAsnGlnIleIleGluGluMetIleLysLysGluAlaIleTyrValAlaTr 737

Qy      2277 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2336
             |||||||||||||||||||||||||||||:::::::::|||   |||||||||:::||
Db       737 pValProAlaHisLysGlyIleGlyGlyAsnGlnGluValAspHisLeuValSerGlnGl 757

Qy      2337 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
             |||||||:::|||||||||||||:::   |||:::
Db       757 yIleArgGlnValLeuPheLeuGluLysIleGlu 768
```

Search completed: June 2, 2005, 05:07:47
Job time : 115.933 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  1, 2005, 11:33:41 ; Search time 168.242 Seconds
                                          (without alignments)
                                          11351.628 Million cell updates/sec

Title:          US-09-610-313B-30
Perfect score:  4583
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2469

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters:      4211384

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155136_15139/app_query.fasta_1.7893
-DB=A_Geneseq_16Dec04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_557_@runat_31052005_155136_15139 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      A_Geneseq_16Dec04:*
                1:  geneseqp1980s:*
                2:  geneseqp1990s:*
                3:  geneseqp2000s:*
                4:  geneseqp2001s:*
                5:  geneseqp2002s:*
                6:  geneseqp2003as:*
                7:  geneseqp2003bs:*
                8:  geneseqp2004s:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3804.5	83.0	1435	8	ADS19486	Ads19486 Modified
2	3782.5	82.5	3183	8	ADP84803	Adp84803 HIV-1 hyb
3	3741.5	81.6	1457	8	ADN36406	Adn36406 Human pro
4	3658	79.8	1003	3	AAB69289	Aab69289 HIV-1 non
5	3649	79.6	999	3	AAB69286	Aab69286 HIV-1 non
6	3644	79.5	1005	3	AAB69287	Aab69287 HIV-1 non
7	3608	78.7	998	6	AAE37601	Aae37601 HIV-1 sub
8	3601	78.6	3025	4	AAB86169	Aab86169 HIV-1 sub
9	3555.5	77.6	1000	3	AAB69282	Aab69282 HIV-1 non
10	3537.5	77.2	1002	2	AAW72993	Aaw72993 HIV isola
11	3537	77.2	854	5	AAM48949	Aam48949 HIV-1 sub
12	3535	77.1	1003	6	AAO30963	Aao30963 HIV pol p
13	3535	77.1	1003	8	ADN36413	Adn36413 HIV prote
14	3532.5	77.1	1002	1	AAP81861	Aap81861 Sequence
15	3532	77.1	1003	6	ABR55489	Abr55489 Amino aci
16	3530	77.0	1003	1	AAP60420	Aap60420 Sequence
17	3530	77.0	1003	1	AAP70861	Aap70861 Sequence
18	3524	76.9	1003	3	AAY70602	Aay70602 Codon opt
19	3524	76.9	1003	3	AAY70601	Aay70601 Corrected
20	3522	76.8	1003	5	AAO19387	Aao19387 Lymphaden
21	3516.5	76.7	1003	2	AAR08060	Aar08060 HIV-1 pol
22	3513	76.7	1015	1	AAP60347	Aap60347 HTLV-III
23	3513	76.7	1015	2	AAR43867	Aar43867 HTLV-III
24	3513	76.7	1015	4	AAB85993	Aab85993 Amino aci
25	3510	76.6	1003	3	AAB10047	Aab10047 HIV-1 pol
26	3510	76.6	1003	3	AAY70600	Aay70600 Wild type
27	3509	76.6	1016	2	AAR08062	Aar08062 AcNPV-HIV
28	3506	76.5	1015	2	AAR43875	Aar43875 HTLV-III
29	3502	76.4	1016	2	AAR08063	Aar08063 HIV-1 pol
30	3499	76.3	2033	2	AAR08056	Aar08056 HIV-1 pol
31	3495	76.3	1016	2	AAR08054	Aar08054 HIV-1 pol
32	3493	76.2	1003	3	AAB69284	Aab69284 HIV-1 non
33	3492	76.2	2033	2	AAR08055	Aar08055 HIV-1 pol
34	3491	76.2	1003	8	ADP20076	Adp20076 Human imm
35	3487	76.1	1003	1	AAP61508	Aap61508 Sequence
36	3487	76.1	1003	2	AAR29705	Aar29705 pol gene
37	3487	76.1	1003	3	AAY77301	Aay77301.HIV-1 (AT
38	3487	76.1	1003	5	AAE35789	Aae35789 ARV-2 (9B
39	3487	76.1	1003	6	ABU63185	Abu63185 Pol prote
40	3487	76.1	1491	1	AAP91048	Aap91048 Transcrip
41	3484	76.0	1003	3	AAB69279	Aab69279 HIV-1 non
42	3482	76.0	739	5	AAU11874	Aau11874 HIV pol p
43	3482	76.0	1003	6	ABU57552	Abu57552 AIDS asso
44	3480	75.9	1012	6	ABU63325	Abu63325 Human lym
45	3477.5	75.9	1003	2	AAR08059	Aar08059 HIV-1 pol

ALIGNMENTS

```
RESULT 1
ADS19486
ID   ADS19486 standard; protein; 1435 AA.
XX
AC   ADS19486;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Modified HIV-1 GagPol protein.
XX
KW   Gene delivery; epithelial cell; respiratory system; mammal;
KW   lentiviral packaging system; expression vector; gagpol; gene expression;
KW   lentivirus; cystic fibrosis transmembrane conductance regulator; CFTR;
KW   cystic fibrosis; CF; deltaPD; transepithelial potential difference;
KW   gene therapy; HIV-1; mutant; mutein.
XX
OS   Human immunodeficiency virus 1.
OS   Synthetic.
XX
PN   US2004037780-A1.
XX
PD   26-FEB-2004.
XX
PF   23-AUG-2002; 2002US-00226638.
XX
PR   19-NOV-2001; 2001AU-00008942.
XX
PA   (PARS/) PARSONS D.
PA   (ANSO/) ANSON D.
PA   (LIMB/) LIMBERIS M.
PA   (FULL/) FULLER M.
XX
PI   Parsons D,  Anson D,  Limberis M,  Fuller M;
XX
```

```
DR   WPI; 2004-191004/18.
DR   N-PSDB; ADS19485.
XX
PT   Delivering exogenous (e.g. CFTR) genes to epithelial cells in the
PT   respiratory tract using a penetration agent and a lentiviral packaging
PT   system, useful for treating e.g. cystic fibrosis.
XX
PS   Disclosure; Fig 16; 45pp; English.
XX
CC   The present invention relates to a method of delivering one or more
CC   exogenous genes for expression in an epithelial cell in the respiratory
CC   system of a mammal to give persistent expression of the gene in the
CC   epithelial cell, using a lentiviral packaging (expression vector) system.
CC   Also disclosed are (i) a recombinant lentiviral packaging system ,
CC   comprising a first nucleic acid molecule including a gag gene sequence
CC   encoding a Gag protein, and a second nucleic acid molecule including a
CC   gagpol gene sequence encoding a GagPol protein (the gagpol gene sequence
CC   has degenerative nucleotide changes in the frame shift sequence AUUUUUU
CC   to reduce the chance of a frame shift which switches expression of the
CC   Gagpol protein to the Gag protein in wild type lentivirus and the
CC   packaging system additionally comprises a lentiviral vector nucleic acid
CC   molecule which does not encode either the gag gene or the gagpol gene or
CC   both, and (ii) a recombinant nucleic acid molecule encoding a lentiviral
CC   gagpol gene having degenerative nucleotide changes in the frame shift
CC   sequence AUUUUUU to reduce the chance of a frame shift which switches
CC   expression of the Gagpol protein to the Gag protein in the wild type
CC   lentivirus. The exogenous gene is expressed in sufficient numbers of
CC   cells and amounts to provide an ameliorating effect for a respiratory
CC   condition. The exogenous gene is cystic fibrosis transmembrane
CC   conductance regulator (CFTR) and the condition is cystic fibrosis
CC   ((CF)the expression of the CFTR gene is sufficient to provide a
CC   significant shift of a reduced deltaPD (transepithelial potential
CC   difference) back to normal levels in the mammal). The cell is non-
CC   terminally differentiated and is capable of differentiating into 2 or
CC   more cell classes e.g. ciliated cells, non-ciliated cells, secretory
CC   cells and basal cells. The exogenous gene may be enzymic. The gagpol gene
CC   sequence has additionally degenerate nucleotide substitutions which
CC   destabilise the hairpin structure associated with the frame shift event.
CC   The gag gene is a truncation of the wild type gagpol gene so that it can
CC   no longer be translated to form Gagpol. The lentivirus is HIV or HIV
CC   derived. The method and recombinant lentiviral packaging system may be
CC   used for delivering nucleic acids to epithelial cells in the respiratory
CC   system of a mammal. In particular they may be used to deliver the CFTR
CC   gene for the treatment of CF. It has been found that the delivery of a
CC   recombinant lentivirus carrying an exogenous gene to the respiratory
CC   system, following the delivery of a non-toxic amount of a penetration
CC   enhancer can provide persistence of expression of a gene product. The
CC   present sequence represents modified HIV-1 GagPol protein.
XX
SQ   Sequence 1435 AA;

Alignment Scores:
Pred. No.:              8.15e-215     Length:        1435
Score:                  3804.50       Matches:       706
Percent Similarity:     94.82%        Conservative:  44
Best Local Similarity:  89.25%        Mismatches:    36
Query Match:            83.01%        Indels:        5
DB:                     8             Gaps:          3

US-09-610-313B-30 (1-2469) x ADS19486 (1-1435)

Qy      13 ATGGCCGAGGCCATGAGCCAGGCCACC---AGCGCCAACATCCTGATGCAGCGCAGCAAC 69
              :::||||||||||||||||||   |||   ||||||   |||:::||||||||   |||
Db     363 LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn 382

Qy      70 TTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCC 129
           |||:::      ::::::   :::||||||||||||||||||||||||||||||||||||
Db     383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla 402

Qy     130 CGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAG 189
           :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422

Qy     190 ATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAG 249
           ||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db     423 MetLysAspCysThrGluArgGlnAlaAsnPheLeuArgGluAspLeuAlaPheProGln 442

Qy     250 GGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAG 309
           ||||||||||:::|||   |||||||||   ||||||||||||||         ||||||
Db     443 GlyLysAlaArgLysPheSerSerGluGlnThrArgAlaAsnSerProIleArgArgGlu 462

Qy     310 CTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGC 363
              ||||||      |||   :::|||      |||||||||||||:::|||||||||
Db     463 ArgGlnValTrpArgArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGly 482

Qy     364 ACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAG 417
           |||:::      :::|||||||||||||||||||||||||||||||||||:::||||||
Db     483 ThrValSerPheSerPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLys 502

Qy     418 GTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAG 477
           :::|||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db     503 IleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGlu 522

Qy     478 GAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATC 537
           |||:::||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db     523 GluMetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIle 542

Qy     538 AAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACC 597
           ||||||||||||||||||||||||   |||||||||||||||   |||||||||||||||
Db     543 LysValArgGlnTyrAspGlnIleProIleGluIleCysGlyHisLysAlaIleGlyThr 562

Qy     598 GTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGC 657
           ||||||:::||||||||||||||||||||||||||||||:::|||||||||:::||||||
Db     563 ValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGly 582

Qy     658 TGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGC 717
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     583 CysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGly 602

Qy     718 ATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACC 777
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     603 MetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuVal 622

Qy     778 GCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCC 837
              ||||||   |||||||||||||||||||||||:::||||||||||||||||||||
Db     623 GluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnPro 642

Qy     838 TACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTG 897
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     643 TyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuVal 662

Qy     898 GACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC 957
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     663 AspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIlePro 682

Qy     958 CACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTAC 1017
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     683 HisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyr 702

Qy    1018 TTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATC 1077
           |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db     703 PheSerValProLeuHisGluAspPheArgLysTyrThrAlaPheThrIleProSerIle 722

Qy    1078 AACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGC 1137
           ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db     723 AsnAsnGluThrProGlyThrArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGly 742

Qy    1138 AGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAAC 1197
           ||||||:::|||||||||||||||||||||   ||||||||||||||||||   :::|||
Db     743 SerProAlaIlePheGlnSerSerMetThrThrIleLeuGluProPheArgLysGlnAsn 762

Qy    1198 CCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATC 1257
           |||::::::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db     763 ProAspLeuValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIle 782

Qy    1258 GGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACC 1317
```

```
              |||||||||||  ||||||||||||||||:::|||||||||||||||||||||
Db        783 GlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThr 802

Qy       1318 ACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCAC 1377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 ThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHis 822

Qy       1378 CCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAAC 1437
              ||||||||||||||||||||||||    |||||||||:::||||||||||||||
Db        823 ProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsn 842

Qy       1438 GACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAG 1497
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||
Db        843 AspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys 862

Qy       1498 GTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTG 1557
              ||||||||||||||||||||||||||||   |||||||||||::::::::||||||
Db        863 ValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeu 882

Qy       1558 ACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCAC 1617
              |||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        883 ThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHis 902

Qy       1618 GGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGAC 1677
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        903 GlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGly 922

Qy       1678 CAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCC 1737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 GlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAla 942

Qy       1738 AAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATC 1797
              :::   |||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 ArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIle 962

Qy       1798 GCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAG 1857
              |||  ||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        963 AlaThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLys 982

Qy       1858 GAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAG 1917
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        983 GluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGlu 1002

Qy       1918 TTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATC 1977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 PheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIle 1022

Qy       1978 GGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCC 2037
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db       1023 GlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAla 1042

Qy       2038 GGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAG 2097
              ||||||||||::::::|||||||||||:::|||||||||||||:::||||||||||||
Db       1043 GlyTyrValThrAsnLysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGln 1062

Qy       2098 AAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATC 2157
              |||||||||||||||||||||   ||||||||||||||||||||||||   ||||||||||
Db       1063 LysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIle 1082

Qy       2158 GTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGC 2217
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db       1083 ValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspArgSerGluSer 1102

Qy       2218 GAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGG 2277
              ||||||||:::|||||||||||||||||||||||||||||||||||||||:::|||
Db       1103 GluLeuValSerGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrp 1122

Qy       2278 GTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGC 2337
              ||||||||||||||||||||||||||||||||||:::||||||||||||||   |||
Db       1123 ValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGly 1142

Qy       2338 ATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              |||||||||||||||||||||||||||||||||
Db       1143 IleArgLysValLeuPheLeuAspGlyIleAsp 1153

RESULT 2
ADP84803
ID   ADP84803 standard; protein; 3183 AA.
XX
AC   ADP84803;
XX
DT   23-SEP-2004  (first entry)
XX
DE   HIV-1 hybrid isolate pNL4-3 protein.
XX
KW   eliciting; inducing; immune response; HIV; antigen; non-pathogenic;
KW   vaccination; vaccine; HIV-I.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO2004056391-A1.
XX
PD   08-JUL-2004.
XX
PF   19-DEC-2003; 2003WO-AU001705.
XX
PR   20-DEC-2002; 2002AU-00953556.
PR   17-SEP-2003; 2003AU-00905067.
XX
PA   (UYNE-) UNIV NEW SOUTH WALES.
XX
PI   Kent SJ,  Purcell DF,  Boyle DB,  Ramsay A,  Thomson S,  Ramshaw IA;
XX
DR   WPI; 2004-500267/47.
DR   N-PSDB; ADP84804.
XX
PT   Eliciting or inducing in a mammal an immune response against HIV-I
PT   subtype AE, B or C by administering sequential doses of a recombinant
PT   plasmid and viral vectors containing the nucleic acid molecules encoding
PT   the HIV antigens.
XX
PS   Disclosure; Fig 12; 280pp; English.
XX
CC   The invention relates to a novel method for eliciting or inducing in a
CC   mammal an immune response directed to a virus, preferably Human
CC   immunodeficiency virus (HIV). The method comprises sequentially
CC   administering to the mammal one or more sequential doses of a recombinant
CC   plasmid vector or recombinant viral vector or its derivative, into which
CC   the nucleic acid molecules encoding all, part or a modified form of two
CC   or more antigens of the virus are incorporated, and one or more optimized
CC   CpG motifs, where the antigens have been rendered substantially non-
CC   pathogenic. The invention further comprises: a method for treating or
CC   preventing viral infection in a mammal; a method of vaccinating a mammal
CC   against a viral pathogen; a method of eliciting or inducing, in a mammal,
CC   an immune response directed to HIV; a vaccine capable of inducing an
CC   immune response directed to a virus comprising the recombinant plasmid
CC   vector or recombinant viral vector or its functional derivative; a
CC   nucleic acid construct or its functional derivative comprising the
CC   plasmid vector or recombinant viral vector; a plasmid vector; a
CC   recombinant viral vector; a pharmaceutical composition comprising the
CC   nucleic acid, plasmid vector or recombinant viral vector constructs. The
CC   method is useful in eliciting or inducing in a mammal an immune response
CC   against HIV-I subtype AE, B or C. This sequence represents the protein of
CC   the HIV-1 hybrid isolate pNL4-3 DNA of the invention.
XX
SQ   Sequence 3183 AA;

Alignment Scores:
Pred. No.:               1.94e-213     Length:        3183
Score:                   3782.50       Matches:       711
Percent Similarity:      87.89%        Conservative:  44
Best Local Similarity:   82.77%        Mismatches:    31
Query Match:             82.53%        Indels:        73
DB:                      8             Gaps:          5

US-09-610-313B-30 (1-2469) x ADP84803 (1-3183)
```

```
Qy     13 ATGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAAC 69
Db    363 LeuAlaGluAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLysGlyAsn 382

Qy     70 TTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCC 129
Db    383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla 402

Qy    130 CGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAG 189
Db    403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422

Qy    190 ATGAAGGACTGCACCGAGCGCCAGGCCAAC------------------------------ 219
Db    423 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys 442

Qy    220 ---------------------------------------------TTCTTCCGCGAGGAC 234
Db    443 GlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPhePheArgGluAsp 462

Qy    235 CTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGC 294
Db    463 LeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSer 482

Qy    294 ------------------------------------------------------------ 294
Db    483 ProProGluGluSerPheArgPheGlyGluGluThrThrThrProSerGlnLysGlnGlu 502

Qy    295 ------------------------------CCCACCAGCCGCGAGCTGCAGGTG------ 318
Db    503 ProIleAspLysGluLeuTyrProLeuAlaProThrArgArgGluLeuGlnValTrpGly 522

Qy    319 CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC------ 372
Db    523 ArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSer 542

Qy    373 ---------------------------------------TTCCCCCAGATCACCCTGTGG 393
Db    543 SerLeuArgSerLeuPheGlySerAspProSerSerGlnPheProGlnIleThrLeuTrp 562

Qy    394 CAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACC 453
Db    563 GlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThr 582

Qy    454 GGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATG 513
Db    583 GlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysProLysMet 602

Qy    514 ATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATC 573
Db    603 IleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIle 622

Qy    574 TGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGC 633
Db    623 CysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGly 642

Qy    634 CGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACC 693
Db    643 ArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThr 662

Qy    694 GTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACC 753
Db    663 ValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThr 682

Qy    754 GAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATC 813
Db    683 GluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIle 702

Qy    814 ACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGAC 873
Db    703 SerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAsp 722

Qy    874 AGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTC 933
Db    723 SerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPhe 742

Qy    934 TGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACC 993
Db    743 TrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSerValThr 762

Qy    994 GTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTAC 1053
Db    763 ValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyr 782

Qy   1054 ACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAAC 1113
Db    783 ThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsn 802

Qy   1114 GTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATC 1173
Db    803 ValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThrLysIle 822

Qy   1174 CTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTG 1233
Db    823 LeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeu 842

Qy   1234 TACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAG 1293
Db    843 TyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGln 862

Qy   1294 CACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTC 1353
Db    863 HisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPhe 882

Qy   1354 CTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCC 1413
Db    883 LeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleValLeuPro 902

Qy   1414 GAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCC 1473
Db    903 GluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAla 922

Qy   1474 AGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAG 1533
Db    923 SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLys 942

Qy   1534 GCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGC 1593
Db    943 AlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArg 962

Qy   1594 GAGATCCTGCGCGAGCCCGTGGACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCC 1653
Db    963 GluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAla 982

Qy   1654 GAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAG 1713
Db    983 GluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLys 1002

Qy   1714 AACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAG 1773
Db   1003 AsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspValLysGln 1022

Qy   1774 CTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCC 1833
Db   1023 LeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThrPro 1042

Qy   1834 AAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAG 1893
Db   1043 LysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrpGln 1062

Qy   1894 GCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTAC 1953
Db   1063 AlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyr 1082

Qy   1954 CAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAAC 2013
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1083 GlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAlaAsn 1102

Qy       2014 CGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTG 2073
              ||||||||||||:::||||||||||||||||||||||||||||||||||||:::|||
Db       1103 ArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValVal 1122

Qy       2074 AGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAG 2133
                    |||||:::|||||||||||||||||||||||||||||    ||||||||||||
Db       1123 ProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGln 1142

Qy       2134 GACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAG 2193
              |||||||||    ||||||||||||||||||||||||||||||||||||||||||||||
Db       1143 AspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGln 1162

Qy       2194 GCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAG 2253
              |||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db       1163 AlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleLys 1182

Qy       2254 AAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAG 2313
              |||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db       1183 LysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGln 1202

Qy       2314 ATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              :::|||||||||||||||    |||||||||||||||||||||||||||||||||||
Db       1203 ValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1221
```

RESULT 3
ADN36406
ID   ADN36406 standard; protein; 1457 AA.
XX
AC   ADN36406;
XX
DT   15-JUL-2004 (first entry)
XX
DE   Human protein for anti-HIV vaccine.
XX
KW   anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX
OS   Homo sapiens.
XX
PN   WO2004035006-A2.
XX
PD   29-APR-2004.
XX
PF   17-OCT-2003; 2003WO-US033112.
XX
PR   18-OCT-2002; 2002US-0419465P.
XX
PA   (AARO-) AARON DIAMOND AIDS RES CENT.
XX
PI   Huang Y, Ho DD, Chen Z;
XX
DR   WPI; 2004-348328/32.
DR   N-PSDB; ADN36405.
XX
PT   Nucleic acid vector comprising at least one HIV sequence operably linked
PT   to a promoter and encoding a protein that does not assemble into viral
PT   particles, useful in immunizing a subject against HIV infection.
XX
PS   Disclosure; SEQ ID NO 20; 166pp; English.
XX
CC   The invention relates to a nucleic acid vector comprising at least one
CC   HIV sequence operably linked to a promoter and encoding a protein that
CC   does not assemble into viral particles. The nucleic acid vector is useful
CC   in immunizing a subject against HIV infection. This sequence corresponds
CC   to a peptide used in the invention.
XX
SQ   Sequence 1457 AA;

```
Alignment Scores:
Pred. No.:            4.08e-211      Length:      1457
Score:                3741.50        Matches:     709
Percent Similarity:   92.24%         Conservative:  28
Best Local Similarity: 88.74%        Mismatches:    44
Query Match:          81.64%         Indels:        19
DB:                   8              Gaps:          3

US-09-610-313B-30 (1-2469) x ADN36406 (1-1457)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||      :::     |||||||||||||||||||||||
Db        383 LeuAlaGluAlaMetSerGlnAla---AsnGlyThrIleLeuMetGlnArgSerAsnPhe 401

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||    ||||||||:::||||||||||||||||||||||||||||||||||||||
Db        402 LysGlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 421

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 441

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||        :::  ||||||    :::
Db        442 LysAspCysThrGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProProHisLysGl 461

Qy        253 AAGGCCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
                ||||||    :::        ||||||      ||||||||||||||||    |||
Db        461 yArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPh 481

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
                          ||||||||||    :::      |||      |||   |||   ||
Db        481 eGlyPheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuTh 501

Qy        373 TTC------------------------------------CCCCAGATCACCCTGTGGCA 395
              |                                      ||||||||||||||||||
Db        501 rSerLeuLysSerLeuPheGlySerAspProLeuSerGlnProGlnIleThrLeuTrpGl 521

Qy        396 GCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGG 455
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db        521 nArgProLeuValSerIleArgValGlyGlyGlnIleLysGluAlaLeuLeu-------- 538

Qy        456 CGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGAT 515
                    ||||||||||||||||||||:::::||||||||||||||||||||||||||||
Db        539 ----AspAspThrValLeuGluGluValAsnLeuProGlyLysTrpLysProLysMetIl 557

Qy        516 CGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTG 575
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db        557 eGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIleGluIleCy 577

Qy        576 CGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCG 635
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        577 sGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyAr 597

Qy        636 CAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGT 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||::
Db        597 gAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrIl 617

Qy        696 GCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGA 755
              :||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        617 eProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrGl 637

Qy        756 GGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCAC 815
              |||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        637 uGluLysIleLysAlaLeuThrAlaIleCysAspGluMetGluLysGluGlyLysIleTh 657

Qy        816 CAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAG 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        657 rLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSe 677

Qy        876 CACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTG 935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 rThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTr 697

Qy        936 GGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGT 995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        697 pGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVa 717

Qy        996 GCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACAC 1055
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        717 lLeuAspValGlyAspAlaTyrPheSerValProLeuTyrGluAspPheArgLysTyrTh 737

Qy       1056 CGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGT 1115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 rAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVa 757

Qy       1116 GCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175
              |||||||||||||||||||||||||||:::|||||||||   ||||||   ||||||||
Db        757 lLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetAlaLysIleLe 777

Qy       1176 GGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTA 1235
              |||||||||||||||:::||||||||||||||||||||||||||||   ||||||||||
Db        777 uGluProPheArgAlaGlnAsnProGluIleValIleTyrGlnTyrGlyAspAspLeuTy 797

Qy       1236 CGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCA 1295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::||
Db        797 rValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHi 817

Qy       1296 CCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCT 1355
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 sLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLe 837

Qy       1356 GTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGA 1415
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        837 uTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGl 857

Qy       1416 GAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAG 1475
              ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 uLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSe 877

Qy       1476 CCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGC 1535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 rGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAl 897

Qy       1536 CCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGA 1595
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        897 aLeuThrAspIleIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGl 917

Qy       1596 GATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGA 1655
              ||||||:::||||||||||||||   ||||||||||||||||||||||||:::|||||
Db        917 uIleLeuLysGluProValHisGlyAlaTyrTyrAspProSerLysAspLeuIleAlaGl 937

Qy       1656 GATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAA 1715
              ||||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db        937 uIleGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAs 957

Qy       1716 CCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCT 1775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 nLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLe 977

Qy       1776 GACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAA 1835
              |||||||||||||||||||||:::|||||||||||||||||||||||||   |||||
Db        977 uThrGluAlaValGlnLysIleSerMetGluSerIleValIleTrpGlyLysIleProLy 997

Qy       1836 GTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGC 1895
              |||||||||||||||   |||||||||||||||||   ||||||   ||||||||||
Db        997 sPheArgLeuProIleProLysGluThrTrpGluThrArgTrpThrAlaTyrTrpGlnAl 1017

Qy       1896 CACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCA 1955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 aThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGl 1037

Qy       1956 GCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCG 2015
              |||||||||:::||||||   |||   |||||||||||||||||||||||||||||||
Db       1037 nLeuGluLysAspProIleAlaGlyValGluThrPheTyrValAspGlyAlaAlaAsnAr 1057

Qy       2016 CGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAG 2075
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db       1057 gGluThrLysMetGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSe 1077

Qy       2076 CCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGA 2135
              ||||||||||||||||||||||||||||||||||||||||   |||||||||||||
Db       1077 rLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleCysLeuAlaLeuGlnAs 1097

Qy       2136 CAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGC 2195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1097 pSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAl 1117

Qy       2196 CCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAA 2255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1117 aGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLy 1137

Qy       2256 GGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGAT 2315
              ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||::
Db       1137 sGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVa 1157

Qy       2316 CGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              :|||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db       1157 lAspLysLeuValSerAsnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1175
```

RESULT 4

```
AAB69289
ID   AAB69289 standard; protein; 1003 AA.
XX
AC   AAB69289;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 94IN476-104 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1003 AA;

Alignment Scores:
Pred. No.:              2.96e-206     Length:         1003
Score:                  3658.00       Matches:        681
Percent Similarity:     98.61%        Conservative:   26
Best Local Similarity:  94.98%        Mismatches:     10
```

```
Query Match:             79.82%      Indels:        0
DB:                      3           Gaps:          0

US-09-610-313B-30 (1-2469) x AAB69289 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::|||||||||||||||||:::|||||||||||||||||:::|||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerLysGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
                 |||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db         21 AlaArgAlaAsnSerProThrSerArgGluLeuGlnValGlnGlyAspAsnProArgSer 40

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGC 399
              |||||||||   |||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GluAlaGlyValGluArgGlnGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg 60

Qy        400 CCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCC 459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ProLeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla 80

Qy        460 GACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGC 519
              ||||||||||||||||||||::::::||||||||||:::|||||||||||||||||||||
Db         81 AspAspThrValLeuGluGluIleAlaLeuProGlyArgTrpLysProLysMetIleGly 100

Qy        520 GGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGC 579
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGly 120

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
              ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||:::
Db        121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsp 140

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 160

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
              ||||||||||||||||||      ||||||:::|||||||||||||||||||||||||||
Db        181 LysIleLysAlaLeuThrGluIleCysLysGluMetGluLysGluGlyLysIleThrLys 200

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
              ||||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db        201 IleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysArgLysAspSerThr 220

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 240

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 260

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
              |||||||||||||||||||||||||||||||||||||    |||    ||||||||||||
Db        261 AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheGlyLysTyrThrAla 280

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 300

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTG 1239
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        321 ProPheArgAlaArgAsnProLysIleValIleTyrGlnTyrMetAspAspLeuTyrVal 340

Qy       1240 GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTG 1299
              ||||||||||||||||||||||   ||||||||||||||||||||||||||   ||||||
Db        341 GlySerAspLeuGluIleGlyHisHisArgAlaLysIleGluGluLeuArgAlaHisLeu 360

Qy       1300 CTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGG 1359
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 380

Qy       1360 ATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAG 1419
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        381 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleLysLeuProGluLys 400

Qy       1420 GAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG 1479
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln 420

Qy       1480 ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTG 1539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440

Qy       1540 ACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATC 1599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIle 460

Qy       1600 CTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATC 1659
              |||:::||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480

Qy       1660 CAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTG 1719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500

Qy       1720 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACC 1779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520

Qy       1780 GAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTC 1839
              ||||||||||||||||||||:::|||||||||||||||||||   |||||||||||||||
Db        521 GluAlaValGlnLysIleAlaIleGluSerIleValIleTrp***LysThrProLysPhe 540

Qy       1840 CGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACC 1899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThr 560

Qy       1900 TGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTG 1959
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpIleProAspTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeu 580

Qy       1960 GAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAG 2019
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600

Qy       2020 ACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTG 2079
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ThrLysValGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSerLeu 620

Qy       2080 ACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGC 2139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 640

Qy       2140 GGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAG 2199
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlyThrGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln 660

Qy       2200 CCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAG 2259
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleAsnLysGlu 680

Qy       2260 AAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGAC 2319
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAsp 700

Qy       2320 AAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
```

```
              :::|||||||||   |||||||||||||||||||||||||||||||||||
Db        701 ArgLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

RESULT 5
AAB69286
ID   AAB69286 standard; protein; 999 AA.
XX
AC   AAB69286;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 96ZM651-8 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 999 AA;

Alignment Scores:
Pred. No.:              9.97e-206     Length:        999
Score:                  3649.00       Matches:       681
Percent Similarity:     98.33%        Conservative:  24
Best Local Similarity:  94.98%        Mismatches:    12
Query Match:            79.62%        Indels:        0
DB:                     3             Gaps:          0

US-09-610-313B-30 (1-2469) x AAB69286 (1-999)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AlaArgAlaAsnSerProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSer 40

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCAGATCACCCTGTGGCAGCGC 399
              |||||||||   ||||||||||||:::|||||||||||||||||||||||||||||||||
Db         41 GluAlaGlyValGluArgGlnGlySerLeuAsnPheProGlnIleThrLeuTrpGlnArg 60

Qy        400 CCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCC 459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ProLeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla 80

Qy        460 GACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGC 519
                 |||||||||||||||:::::||||||||||||||||||||||||||||||||||||||
Db         81 GlyAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGly 100

Qy        520 GGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGC 579
              ||||||||||||||||:::||||||||||||||||||||     :::|||||||||||||
Db        101 GlyIleGlyGlyPheIleGluValArgGlnTyrAspGlnIleProMetGluIleCysGly 120

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 140

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 160

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LysIleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLys 200

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThr 220

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 240

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 260

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
              ||||||||||||||||||||||||||||||||||||||   |||||||||||||||||||
Db        261 AspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThrAla 280

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
              ||||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db        281 PheThrIleProSerThrAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 300

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTG 1239
              ||||||||||:::||||||:::||||||||||||||||||||||||||||||||||||||
Db        321 ProPheArgAlaGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrVal 340

Qy       1240 GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTG 1299
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        341 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeu 360

Qy       1300 CTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGG 1359
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 380

Qy       1360 ATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAG 1419
              |||||||||||||||||||||||||||||||||||||||||||:::|||   ||||||
Db        381 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuAlaGluLys 400

Qy       1420 GAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG 1479
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln 420

Qy       1480 ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTG 1539
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 IleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440

Qy       1540 ACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATC 1599
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnLysGluIle 460

Qy       1600 CTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATC 1659
              |||:::||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db        461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480

Qy       1660 CAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTG 1719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500

Qy       1720 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACC 1779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520

Qy       1780 GAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTC 1839
              ||||||||||||||||||||:::||||||||||||||||||||||||   |||||||||
Db        521 GluAlaValGlnLysIleAlaLeuGluSerIleValIleTrpGlyLysIleProLysPhe 540

Qy       1840 CGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACC 1899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThr 560

Qy       1900 TGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTG 1959
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        561 TrpIleProGluTrpGluPheValAsnThrProLeuLeuValLysLeuTrpTyrGlnLeu 580

Qy       1960 GAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAG 2019
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600

Qy       2020 ACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTG 2079
              ||||||:::||||||||||||||:::||||||||||||||||||||||||||:::|||
Db        601 ThrLysLeuGlyLysAlaGlyTyrIleThrAspArgGlyArgGlnLysIleValThrLeu 620

Qy       2080 ACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGC 2139
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSer 640

Qy       2140 GGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAG 2199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaHis 660

Qy       2200 CCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAG 2259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlu 680

Qy       2260 AAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGAC 2319
              :::||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAsp 700

Qy       2320 AAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 LysLeuValSerLysGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717
```

RESULT 6
AAB69287

```
ID   AAB69287 standard; protein; 1005 AA.
XX
AC   AAB69287;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 96ZM751-3 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1005 AA;

Alignment Scores:
Pred. No.:              1.96e-205     Length:         1005
Score:                  3644.00       Matches:        685
Percent Similarity:     97.23%        Conservative:   18
Best Local Similarity:  94.74%        Mismatches:     14
Query Match:            79.51%        Indels:         6
DB:                     3             Gaps:           1

US-09-610-313B-30 (1-2469) x AAB69287 (1-1005)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::||||||||||||:::|||:::|||   |||    |||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGluGlyGluAlaGlyGluLeuProSerGluGln 20

Qy        280 AACCGCGCC-----------------AACAGCCCCACCAGCCGCGAGCTGCAGGTGCGC 321
                 |||||                 |||||||||||||||||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrSerSerAsnSerProThrSerArgGluLeuGlnValArg 40

Qy        322 GGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAG 381
              |||||||||||     ||||||||||||||||||||||||||||||||||   ||||||
Db         41 GlyAspAsnProCysProGluAlaGlyAlaGluArgGlnGlyThrLeuAsnCysProGln 60

Qy        382 ATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCC 441
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIleLysGluAla 80

Qy        442 CTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGG 501
              ||||||||||||||||||||||||||||||||||||||||:::::|||||||||||||||
Db         81 LeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrp 100

Qy        502 AAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATC 561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIle 120

Qy        562 CTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTG 621
              ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        121 LeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVal 140

Qy        622 AACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGC 681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSer 160
```

```
Qy        682 CCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAG 741
              |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        161 ProIleGluThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGln 180

Qy        742 TGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAG 801
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLys 200

Qy        802 GAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATC 861
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIle 220

Qy        862 AAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGC 921
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArg 240

Qy        922 ACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAG 981
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLys 260

Qy        982 AAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGAC 1041
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 LysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGly 280

Qy       1042 TTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGC 1101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArg 300

Qy       1102 TACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGC 1161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSer 320

Qy       1162 ATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTAC 1221
              |||   |||||||||||||||||||||     :::|||||||||||||||||||||||||
Db        321 MetIleLysIleLeuGluProPheArgThrGlnAsnProGluIleValIleTyrGlnTyr 340

Qy       1222 ATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAG 1281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 MetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGlu 360

Qy       1282 GAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAG 1341
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLys 380

Qy       1342 GAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCC 1401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 GluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnPro 400

Qy       1402 ATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAG 1461
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLys 420

Qy       1462 CTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTG 1521
              |||||||||   ||||||||||||   |||||||||||||||||||||||||||||||||
Db        421 LeuAsnTrp***SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeu 440

Qy       1522 CGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTG 1581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeu 460

Qy       1582 GCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAG 1641
              ||||||:::||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        461 AlaGluSerArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLys 480

Qy       1642 GACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAG 1701
              ||||||:::||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        481 AspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnValTyrGln 500

Qy       1702 GAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAAC 1761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsn 520

Qy       1762 GACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGG 1821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 AspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrp 540

Qy       1822 GGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACC 1881
              ||||||      ||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThr 560

Qy       1882 GACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVal 580

Qy       1942 AAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGAC 2001
              ||||||||||||||||||||||||||||||      ||||||||||||:::|||||||||
Db        581 LysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrTyrTyrValAsp 600

Qy       2002 GGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGG 2061
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArg 620

Qy       2062 CAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAG 2121
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        621 GlnLysIleValThrLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGln 640

Qy       2122 CTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTG 2181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeu 660

Qy       2182 GGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAG 2241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGlu 680

Qy       2242 CAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGC 2301
              ||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        681 GlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGly 700

Qy       2302 GGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGAC 2361
              ||||||||||||:::|||||||||||||   |||||||||||||||||||||||||||||
Db        701 GlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAsp 720

Qy       2362 GGCATCGAT 2370
              |||||||||
Db        721 GlyIleAsp 723
```

RESULT 7
AAE37601

```
ID   AAE37601 standard; protein; 998 AA.
XX
AC   AAE37601;
XX
DT   23-OCT-2003  (revised)
DT   27-AUG-2003  (first entry)
XX
DE   HIV-1 subtype C isolate Du151 reverse transcriptase (RT) protein.
XX
KW   Regulatory gene; accessory gene; HIV; human immunodeficiency virus;
KW   vaccine; infection; gene therapy; reverse transcriptase; RT; enzyme.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO2003037919-A2.
XX
PD   08-MAY-2003.
XX
PF   31-OCT-2002; 2002WO-IB004550.
XX
PR   31-OCT-2001; 2001ZA-00008978.
XX
PA   (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.
PA   (UYCA-) UNIV CAPE TOWN.
XX
PI   Williamson C,  Van Harmelen JH,  Gray CM,  Bourn W,  Karim SA;
XX
```

```
DR   WPI; 2003-430497/40.
DR   N-PSDB; AAD29258.
XX
PT   New molecules comprising HIV-1 subtype isolate regulatory/accessory
PT   genes, useful for manufacturing a vaccine for treating or preventing HIV
PT   infection.
XX
PS   Disclosure; Page 94-97; 97pp; English.
XX
CC   The invention relates to molecules comprising HIV-1 subtype isolate
CC   regulatory/accessory genes (tat, nef and rev genes) and modifications and
CC   derivatives thereof. The invention also provides proteins encoded by such
CC   genes. Sequences of the invention are useful for manufacturing vaccines
CC   for treating or preventing human immunodeficiency virus (HIV) infections.
CC   They are also useful in gene therapy. The present sequence is HIV-1
CC   subtype C isolate Du151 reverse transcriptase (RT) protein. (Updated on
CC   23-OCT-2003 to standardise OS field)
XX
SQ   Sequence 998 AA;

Alignment Scores:
Pred. No.:              2.54e-203     Length:         998
Score:                  3608.00       Matches:        670
Percent Similarity:     97.63%        Conservative:   29
Best Local Similarity:  93.58%        Mismatches:     17
Query Match:            78.73%        Indels:         0
DB:                     6             Gaps:           0

US-09-610-313B-30 (1-2469) x AAE37601 (1-998)

Qy        223 TTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAAC 282
              |||||||||:::||||||||||||||||||:::||||||||||||||||||||||||   
Db          1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20

Qy        283 CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAG 342
              |||||||||||||||||||||||||||||||||||||||   :::||||||||||||||||
Db         21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu 40

Qy        343 GCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
                 |||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro 60

Qy        403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
              |||||||||||||||:::|||||||||   :::||||||||||||||||||||||||||
Db         61 LeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp 80

Qy        463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
              |||||||||||||||:::::::::|||||||||||||||||||||||||||||||||||
Db         81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 100

Qy        523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 IleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLys 120

Qy        583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140

Qy        643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160

Qy        703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180

Qy        763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200

Qy        823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        201 GlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220

Qy        883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240

Qy        943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp 260

Qy       1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        261 ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 280

Qy       1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 300

Qy       1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
              |||||||||||||||||||||:::||||||||||   ||||||||||||||||||||||||
Db        301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320

Qy       1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGC 1242
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 340

Qy       1243 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1302
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeu 360

Qy       1303 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATG 1362
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380

Qy       1363 GGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1422
              ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||:::
Db        381 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400

Qy       1423 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1482
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420

Qy       1483 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1542
              |||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440

Qy       1543 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1602
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 AspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu 460

Qy       1603 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1662
              :::|||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db        461 LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln 480

Qy       1663 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1722
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 500

Qy       1723 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1782
              ||||||||||||||||||   ||||||   ||||||||||||||||||||||||||||||
Db        501 ThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThrGlu 520

Qy       1783 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1842
              |||||||||||||||::::::|||||||||||||||||||||||||||||||||||||||
Db        521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540

Qy       1843 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1902
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        541 LeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThrTrp 560

Qy       1903 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 580
```

```
Qy       1963 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2022
              |||||||||||    ||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 600

Qy       2023 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2082
              |||||||||||||||||||||||||||||||||||||||||||||||||:::|||:::
Db        601 LysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeuSer 620

Qy       2083 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2142
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640

Qy       2143 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2202
              |||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        641 SerGluValAsnIleIleThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 660

Qy       2203 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2262
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        661 AspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluArg 680

Qy       2263 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2322
              ||||||||||||||||||||||||||   |||:::  ||||||||    :::||||||
Db        681 ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700

Qy       2323 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              ||||||||:::||||||||||||||||||    |||||||||||||||
Db        701 LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716
```

RESULT 8
AAB86169
ID   AAB86169 standard; protein; 3025 AA.
XX
AC   AAB86169;
XX
DT   11-SEP-2003 (revised)
DT   21-AUG-2001 (first entry)
XX
DE   HIV-1 subtype C protein fragment #1.
XX
KW   Infection; diagnosis; human; humoral immune response; antiviral;
KW   cellular immune response; vaccine; treatment; gene therapy.
XX
OS   Human immunodeficiency virus 1.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference   1. .3025
FT   /label= Xaa
FT   /note= "Xaa represents a stop codon"
XX
PN   DE10056747-A1.
XX
PD   31-MAY-2001.
XX
PF   16-NOV-2000; 2000DE-01056747.
XX
PR   16-NOV-1999;   99DE-01055089.
XX
PA   (SHAO/) SHAO Y.
PA   (GENE-) GENEART GMBH GES ANGEWANDTE BIOTECHNOLOG.
XX
PI   Wagner R, Wolf H, Shao Y, Graf M;
XX
DR   WPI; 2001-336417/36.
DR   N-PSDB; AAH20870.
XX
PT   New nucleic acid sequences from a human immune deficiency virus
PT   intersubtype, useful for treatment, prevention and diagnosis of
PT   infection.
XX
PS   Disclosure; Fig 8A-O; 48pp; German.
XX
CC   This invention describes a novel polynucleotide isolated from human
CC   immunodeficiency virus type 1 subtype C/B' which can be used for the
CC   induction of specific humoral and cellular immune responses. (I) and
CC   polypeptides (II) encoded by them, are useful in pharmaceuticals,
CC   vaccines and diagnostic agents, particularly for treatment or prevention
CC   of human immune deficiency virus-1 (HIV-1) infections, also for rational
CC   design of test or therapeutic reagents, or gene therapy vectors.
CC   Polypeptides, especially antibodies, specifically directed against (II)
CC   are similarly useful as pharmaceutical and diagnostic agents. (I) are
CC   specific for intersubtype C/B' of HIV-1 so are useful in regions
CC   (particularly China and South-East Asia) where this subtype is prevalent.
CC   The products of the invention have antiviral activity. This sequence
CC   represents a protein encoded by the HIV-1 subtype C genome described in
CC   the method of the invention. (Updated on 11-SEP-2003 to standardise OS
CC   field)
XX
SQ   Sequence 3025 AA;

```
Alignment Scores:
Pred. No.:              8.62e-203     Length:         3025
Score:                  3601.00       Matches:        686
Percent Similarity:     92.15%        Conservative:   42
Best Local Similarity:  86.84%        Mismatches:     55
Query Match:            78.57%        Indels:         7
DB:                     4             Gaps:           4

US-09-610-313B-30 (1-2469) x AAB86169 (1-3025)

Qy         15 GGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAA 74
              |||||||||:::||||||         |||       |||   ||||||    :::|||
Db        416 GlyArgGlyAsnGluProAsnLysGln---CysHisThrAspAlaGluLysGlnPhe*** 434

Qy         75 GGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAA 134
                       :::      :::   ||||||  ||||||       |||       :::
Db        435 ArgLeu***LysAsnCys***MetPheGlnLeuTrpGlnGlyArgAlaHisSerGlnLys 454

Qy        135 CTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAA 194
              |||       |||    ::::::|||||||||||:::    :::         |||:::|||
Db        455 LeuGlnGlyPro***GluLysGlyLeuLeuGluMetTrpLysArgArgThrProAsnGlu 474

Qy        195 GGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAA 254
                 |||:::         |||||| |||||||||||:::||||||   |||||||||::
Db        475 ArgLeuTyr***GluThrGlyGln-PhePheArgGluAsnLeuAlaLeuProGlnGlyAr 494

Qy        255 GGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCA 314
              :||||||||||||   |||||||||   ||||||||||||||||||      ||||||||
Db        494 gAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerProThrArgGlyGluLeuGl 514

Qy        315 GGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCT 368
              ||||      |||   :::|||      |||||||||||||||   ||||||||||||::
Db        514 nValTrpGlyArgAspAsnAsnSerIleSerGluAlaGlyAlaAsnArgGlnGlyThrIl 534

Qy        369 G------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGG 422
              :      |||||||||||||||||||||||||||||||||||||:::||||||:::||
Db        534 eSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGl 554

Qy        423 CGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGAT 482
              |||||||:::||||||||||||||||||||||||||   |||||||||||||||:::::
Db        554 yGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaGlyAspThrValLeuGluAspLe 574

Qy        483 GAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGT 542
              ::::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 uAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysVa 594

Qy        543 GCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCT 602
              |||||||||||:::||||||   ||||||||||||||||||   ||||||||||||||||||
Db        594 lArgGlnTyrGluGlnIleProIleGluIleCysGlyHisLysAlaIleGlyThrValLe 614

Qy        603 GATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCAC 662
              |:::||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        614 uValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysTh 634

Qy        663 CCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGA 722
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 rLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAs 654
```

```
Qy        723 CGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCAT 782
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        654 pGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIl 674

Qy        783 CTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAA 842
              ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        674 eCysAspGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAs 694

Qy        843 CACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTT 902
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 nThrProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPh 714

Qy        903 CCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCC 962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        714 eArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPr 734

Qy        963 CGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAG 1022
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        734 oAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSe 754

Qy       1023 CGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAA 1082
              |:::|||||   ||||||||||||||||||||||||||||||||||||||   |||||
Db        754 rIleProLeuTyrGluAspPheArgLysTyrThrAlaPheThrIleProSerArgAsnAs 774

Qy       1083 CGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCC 1142
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 nGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerLe 794

Qy       1143 CAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGA 1202
               :::||||||||||||||||||||||||||||||||||||||||   :::||||||
Db        794 uAlaIlePheGlnSerSerMetThrLysThrLeuGluProPheArgLysGlnAsnProGl 814

Qy       1203 GATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCA 1262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 yIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGl 834

Qy       1263 GCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCC 1322
              ||||||   ||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        834 nHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrPr 854

Qy       1323 CGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGA 1382
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        854 oAspLysLysHis---LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAs 873

Qy       1383 CAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACAT 1442
              ||||||||||||||||   :::||||||||||:::|||||||||||||||||||||||||
Db        873 pLysTrpThrValGlnProThrGlnLeuProGluLysAspSerTrpThrValAsnAspIl 893

Qy       1443 CCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCG 1502
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        893 eGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValAr 913

Qy       1503 CCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGA 1562
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 gGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGl 933

Qy       1563 GGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGT 1622
              ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        933 uGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyVa 953

Qy       1623 GTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTG 1682
              ||||||||||||||||||||||||:::||||||||||||||||||||   :::|||||
Db        953 lTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGluGlnTr 973

Qy       1683 GACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGAT 1742
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 pThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMe 993

Qy       1743 GCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCAT 1802
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        993 tArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMe 1013

Qy       1803 GGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGAC 1862
              ||||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1013 tGluGlyIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluTh 1033

Qy       1863 CTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGT 1922
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 rTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVa 1053

Qy       1923 GAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGC 1982
              ||||||||||||||||||||||||||||||||||||||||:::||||||:::|||
Db       1053 lAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysAspProIleValGlyVa 1073

Qy       1983 CGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTA 2042
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1073 lGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTy 1093

Qy       2043 CGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGAC 2102
              ||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db       1093 rValThrAspArgGlyArgLysLysIleValSerLeuThrGluThrThrAsnGlnLysTh 1113

Qy       2103 CGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGAC 2162
              |||||||||||||||   :::|||||||||||||||||||||||||||||||||||||||
Db       1113 rGluLeuGlnAlaIleCysIleAlaLeuGlnAspSerGlySerGluValAsnIleValTh 1133

Qy       2163 CGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCT 2222
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1133 rAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLe 1153

Qy       2223 GGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCC 2282
              ||||||||||||||||||||||||:::|||||||||:::|||||||||||||||||||||
Db       1153 uValAsnGlnIleIleGluGlnLeuMetLysLysGluArgValTyrLeuSerTrpValPr 1173

Qy       2283 CGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCG 2342
              ||||||||||||||||||||||||||||||:::|||||||||||||||   |||||||||
Db       1173 oAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleAr 1193

Qy       2343 CAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              ||||||||||||||||||||||||||||
Db       1193 gLysValLeuPheLeuAspGlyIleAsp 1202
```

RESULT 9
AAB69282

```
ID   AAB69282 standard; protein; 1000 AA.
XX
AC   AAB69282;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 92RW009-6 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
```

```
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1000 AA;

Alignment Scores:
Pred. No.:              3.07e-200     Length:         1000
Score:                  3555.50       Matches:        664
Percent Similarity:     96.52%        Conservative:   29
Best Local Similarity:  92.48%        Mismatches:     24
Query Match:            77.58%        Indels:         1
DB:                     3             Gaps:           1

US-09-610-313B-30 (1-2469) x AAB69282 (1-1000)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db      1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArgLysPheSerProGluGln 20

Qy    280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
Db     21 ThrGlyAlaAsnSerProThrSerArgGluLeuTrpAsnGlyGlyArgAspSerLeuSer 40

Qy    337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAG 396
Db     41 SerGluThrGlyAlaGluArgGlnGlyThrPheAsnPheProGlnIleThrLeuTrpGln 60

Qy    397 CGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGC 456
Db     61 ArgProLeuValThrValLysIleGlyGlyGlnLeuArgGluAlaLeuLeuAspThrGly 80

Qy    457 GCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATC 516
Db     81 AlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIle 100

Qy    517 GGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGC 576
Db    101 GlyGlyIleGlyGlyPheIleLysValLysGlnTyrAspGlnIleLeuIleGluIleCys 120

Qy    577 GGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGC 636
Db    121 GlyLysLysAlaIleGlyThrValLeuValGlyProThrSerValAsnIleIleGlyArg 140

Qy    637 AACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTG 696
Db    141 AsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal 160

Qy    697 CCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAG 756
Db    161 ProValAlaLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 180

Qy    757 GAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACC 816
Db    181 GluLysIleLysAlaLeuArgGluIleCysThrGluMetGluLysGluGlyLysIleSer 200

Qy    817 AAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGC 876
Db    201 LysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 220

Qy    877 ACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936
Db    221 ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp 240

Qy    937 GAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTG 996
Db    241 GluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVal 260

Qy    997 CTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACC 1056
Db    261 LeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThr 280

Qy   1057 GCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTG 1116
Db    281 AlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 300

Qy   1117 CTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTG 1176
Db    301 LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAsnSerMetThrLysIleLeu 320

Qy   1177 GAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTAC 1236
Db    321 GluProPheArgAlaGlnAsnGlnGluIleValIleTyrGlnTyrMetAspAspLeuTyr 340

Qy   1237 GTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCAC 1296
Db    341 ValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHis 360

Qy   1297 CTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG 1356
Db    361 LeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 380

Qy   1357 TGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAG 1416
Db    381 TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGlu 400

Qy   1417 AAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGC 1476
Db    401 LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 420

Qy   1477 CAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCC 1536
Db    421 GlnIleTyrProGlyValLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAla 440

Qy   1537 CTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAG 1596
Db    441 LeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 460

Qy   1597 ATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAG 1656
Db    461 IleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGlu 480

Qy   1657 ATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAAC 1716
Db    481 IleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 500

Qy   1717 CTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTG 1776
Db    501 LeuLysThrGlyLysTyrAlaLysArgArgThrAlaHisThrAsnAspValLysGlnLeu 520

Qy   1777 ACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAG 1836
Db    521 ThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrpGlyLysThrProLys 540

Qy   1837 TTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCC 1896
Db    541 PheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAla 560

Qy   1897 ACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAG 1956
Db    561 ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 580

Qy   1957 CTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGC 2016
Db    581 LeuGluLysGluProIleLeuGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg 600

Qy   2017 GAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGC 2076
```

```
Db        601 GluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSer 620

Qy       2077 CTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGAC 2136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 LeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAsp 640

Qy       2137 AGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCC 2196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 SerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 660

Qy       2197 CAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAG 2256
              |||||||||   ||||||||||||   |||||||||||||||||||||||||||||||||
Db        661 GlnProAspSerSerGluSerGluAlaValAsnGlnIleIleGluGlnLeuIleLysLys 680

Qy       2257 GAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATC 2316
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        681 GluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 700

Qy       2317 GACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              |||||||||||||||   |||||||||:::|||||||||||||||||||||||
Db        701 AspLysLeuValSerSerGlyIleArgArgValLeuPheLeuAspGlyIleAsp 718

RESULT 10
AAW72993
ID   AAW72993 standard; protein; 1002 AA.
XX
AC   AAW72993;
XX
DT   25-MAR-2003  (revised)
DT   15-FEB-1999  (first entry)
XX
DE   HIV isolate LAV.MAL pol protein.
XX
KW   LAV.MAL; HIV; human immunodeficiency virus; diagnosis; vaccine; AIDS;
KW   pol.
XX
OS   Human T-lymphotropic virus.
XX
PN   US5824482-A.
XX
PD   20-OCT-1998.
XX
PF   06-JUN-1995;   95US-00471474.
XX
PR   23-JUN-1986;   86FR-00040138.
PR   13-APR-1987;   87US-00038330.
PR   19-FEB-1991;   91US-00656797.
PR   10-DEC-1992;   92US-00988530.
PR   18-NOV-1993;   93US-00154397.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Wain-Hobson S,  Sonigo P,  Alizon M,  Montagnier L;
XX
DR   WPI; 1998-582548/49.
DR   N-PSDB; AAV63467.
XX
PT   Human immunodeficiency virus isolate LAV(MAL) - and method for detecting
PT   anti-HIV antibodies.
XX
PS   Disclosure; Fig 7B-E; 47pp; English.
XX
CC   This is the amino acid sequence of the pol protein of lymphadenopathy
CC   associated virus LAV.MAL (CNCM I-641), a new virus isolate from Zaire
CC   that is responsible for diseases clinically related to AIDS. The sequence
CC   was deduced from an open reading frame (ORF) of the LAV.MAL genome (see
CC   AAV63467). 7 ORFs (see AAW72992-98) were identified. Specific peptides of
CC   the envelope glycoprotein can be used as antigens in a claimed method for
CC   the in vitro detection of an antibody directed against LAV. The method is
CC   useful for the diagnosis of AIDS or pre-AIDS, or to detect antibodies in
CC   patients, asymptomatic carriers and in blood-related products. LAV viral
CC   antigens are also useful in vaccines. (Updated on 25-MAR-2003 to correct
CC   PR field.)
XX
SQ   Sequence 1002 AA;

Alignment Scores:
Pred. No.:               3.49e-199      Length:        1002
Score:                   3537.50        Matches:       657
Percent Similarity:      96.67%         Conservative:  39
Best Local Similarity:   91.25%         Mismatches:    21
Query Match:             77.19%         Indels:        3
DB:                      2              Gaps:          2

US-09-610-313B-30 (1-2469) x AAW72993 (1-1002)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
                 |||||||||||||||||||||||||||:::|||         ||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACCCTG 390
              ||||||   |||||||||||||||||         :::   :::||||||||||||||||
Db         41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPheSerPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||||:::|||:::::::::|||||||||||||:::|||||||||||||||||
Db         61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              ||||||||||||||||||||||||:::|||:::|||||||||||||||||||||||||||
Db        121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              |||||||||||||||||||||||||||   ||||||::::::||||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300
```

```
Qy     1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
            ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db      301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy     1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230
            ||||||||||||||||||   :::||||||||||||||||||||||||||||||||||||
Db      321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy     1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290
            |||||||||||||||||||||||||||||||||||||||      |||||||||||||||
Db      341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy     1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350
            :::|||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db      361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy     1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1410
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db      381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400

Qy     1411 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1470
            |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy     1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530
            ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db      421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440

Qy     1531 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1590
            ||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db      441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460

Qy     1591 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1650
            ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||:::
Db      461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy     1651 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1710
            |||||||||||||||||||||      |||||||||||||||||||||||||||   :::
Db      481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500

Qy     1711 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1770
            |||||||||||||||||||||||||||::::::::::::|||||||||||||||||||||
Db      501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520

Qy     1771 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1830
            ||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db      521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540

Qy     1831 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890
            |||||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db      541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy     1891 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1950
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy     1951 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2010
            ||||||||||||   |||||||||:::|||||||||||||||||||||||||||||||||
Db      581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy     2011 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2070
            |||||||||||||||   |||||||||||||||||||||||||||||||||||||||:::
Db      601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy     2071 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
            ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db      621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy     2131 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2190
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy     2191 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2250
            ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db      661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

Qy     2251 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2310
            :::|||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db      681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy     2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
            |||:::|||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db      701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

```
RESULT 11
AAM48949
ID   AAM48949 standard; protein; 854 AA.
XX
AC   AAM48949;
XX
DT   29-AUG-2003  (revised)
DT   19-APR-2002  (first entry)
XX
DE   HIV-1 subtype C isolate Du151 pol protein.
XX
KW   HIV-1 subtype C; vaccine; HIV infection; AIDS; pol; antiviral.
XX
OS   Human immunodeficiency virus; type I.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 58
FT                   /note= "encoded by TGA"
FT   Misc-difference 72
FT                   /note= "encoded by TAA"
FT   Misc-difference 114
FT                   /note= "encoded by TAA"
FT   Misc-difference 117
FT                   /note= "encoded by TAA"
XX
PN   WO200204494-A2.
XX
PD   17-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-IB001208.
XX
PR   07-JUL-2000; 2000US-0216995P.
PR   10-JUL-2000; 2000ZA-00003437.
PR   15-SEP-2000; 2000ZA-00004924.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
PA   (UYCA-) UNIV CAPE TOWN.
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Williamson C,  Swanstrom RI,  Morris L,  Karim SA,  Johnston RE;
XX
DR   WPI; 2002-171700/22.
DR   N-PSDB; AAL41592.
XX
PT   Selecting HIV-1 subtype C isolates, which are useful in developing
PT   vaccines against HIV infection, comprises isolating viruses with high
PT   sequence identity to a consensus sequence whose phenotype is associated
PT   with the HIV subtype.
XX
PS   Claim 19; Page 67; 69pp; English.
XX
CC   The present invention relates to a process for the selection of human
CC   immunodeficiency virus (HIV) subtype isolates for use in the development
CC   of a prophylactic and/or therapeutic pharmaceutical composition. The
CC   process involves selecting isolated virus or viruses with a high sequence
CC   identity to a consensus sequence and a phenotype which is associated with
CC   transmission for the particular HIV subtype. The composition can be used
CC   in the production of vaccines against HIV. The present sequence is the
CC   HIV-1 subtype C isolate Du151 pol protein. (Updated on 29-AUG-2003 to
CC   standardise OS field)
XX
SQ   Sequence 854 AA;
```

```
Alignment Scores:
Pred. No.:               3.59e-199     Length:        854
Score:                   3537.00       Matches:       667
Percent Similarity:      94.44%        Conservative:  30
Best Local Similarity:   90.38%        Mismatches:    30
Query Match:             77.18%        Indels:        11
DB:                      5             Gaps:          2

US-09-610-313B-30 (1-2469) x AAM48949 (1-854)

Qy      171 CGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTT------ 224
Db      118 ArgGlnGlyPheProSerHisAspValValLysArgArgProValProSerLeuHisAla 137

Qy      225 --------------------CCGCGAGGACCTGGCC---TTCCCCCAGGGCAAGGCCCG 260
Db      138 CysArgSerThrLeuGluAspProArgValPro-SerSerPheProGlnGlyProAlaAr 157

Qy      261 CGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCG 320
Db      157 gGlnPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeuGlnValAr 177

Qy      321 CGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCA 380
Db      177 gArgAspAsnProArgSerGluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGl 197

Qy      381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db      197 nIleThrLeuTrpGlnArgProLeuValSerIleLysIleGlyGlyGlnThrArgGluAl 217

Qy      441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db      217 aLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTr 237

Qy      501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db      237 pLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIl 257

Qy      561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db      257 eLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVa 277

Qy      621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db      277 lAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSe 297

Qy      681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db      297 rProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGl 317

Qy      741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db      317 nTrpProLeuThrGluValLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLy 337

Qy      801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db      337 sGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIlePheAlaIl 357

Qy      861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db      357 eLysLysGluAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysAr 377

Qy      921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db      377 gThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLy 397

Qy      981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db      397 sLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGl 417

Qy     1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db      417 yPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleAr 437

Qy     1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db      437 gTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSe 457

Qy     1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTA 1220
Db      457 rMetThrLysIleLeuGluProPheArgAlaLysAsnProGluIleValIleTyrGlnTy 477

Qy     1221 CATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1280
Db      477 rMetAlaAlaLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGl 497

Qy     1281 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1340
Db      497 uGluLeuArgGluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLy 517

Qy     1341 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1400
Db      517 sGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnPr 537

Qy     1401 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1460
Db      537 oIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLy 557

Qy     1461 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1520
Db      557 sLeuAsnTrpThrSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLe 577

Qy     1521 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1580
Db      577 uArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLe 597

Qy     1581 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1640
Db      597 uAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLy 617

Qy     1641 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1700
Db      617 sAspLeuIleAlaGluIleGlnLysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGl 637

Qy     1701 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1760
Db      637 nGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysArgArgThrThrHisThrAs 657

Qy     1761 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1820
Db      657 nAspValLysGlnLeuThrGluAlaValGlnLysIleSerLeuGluSerIleValThrTr 677

Qy     1821 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1880
Db      677 pGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpTh 697

Qy     1881 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1940
Db      697 rAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVa 717

Qy     1941 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 2000
Db      717 lLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPheTyrValAs 737

Qy     2001 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2060
Db      737 pGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyAr 757

Qy     2061 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2120
Db      757 gGlnLysIleValThrLeuSerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGl 777

Qy     2121 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2180
Db      777 nLeuAlaLeuGlnAspSerGluSerGluValAsnIleValThrAspSerGlnTyrAlaLe 797
```

```
Qy       2181 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2240
              ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        797 uGlyIleIleGlnAlaGlnProAspArgSerGluSerGluLeuValAsnGlnIleIleGl 817

Qy       2241 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2300
              ||||||||||||||||||:::   |||||||||||||||||||||||||||||||||||
Db        817 uGlnLeuIleLysLysGluArgAlaTyrLeuSerTrpValProAlaHisLysGlyIleGl 837

Qy       2301 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2352
              ||||:::||||||:::||||||||||||||   ||||||||||||||||
Db        837 yGlyAspGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 854
```

RESULT 12
AAO30963
ID   AAO30963 standard; protein; 1003 AA.
XX
AC   AAO30963;
XX
DT   06-OCT-2003 (first entry)
XX
DE   HIV pol protein.
XX
KW   Human immunodeficiency virus; HIV; pol; IV9; immunotherapy; infection;
KW   epitope; vaccine; enzyme.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003051285-A2.
XX
PD   26-JUN-2003.
XX
PF   29-OCT-2002; 2002WO-US034620.
XX
PR   29-OCT-2001; 2001US-0345116P.
XX
PA   (GENZ) GENZYME CORP.
PA   (GEHO) GEN HOSPITAL CORP DBA MASSACHUSETTS GEN.
XX
PI   Nicolette CA, Walker BD;
XX
DR   WPI; 2003-541603/51.
XX
PT   Novel synthetic therapeutic peptide compounds useful for inducing an
PT   immune response in a mammal, specifically cross-react with HIV IV9
PT   epitopes.
XX
PS   Claim 8; Page 65-67; 67pp; English.
XX
CC   The invention relates to synthetic therapeutic peptide compounds that
CC   specifically cross-react with HIV (human immunodeficiency virus) pol
CC   protein IV9 epitopes. The invention is useful for treating HIV infection.
CC   The invention is also used in immunotherapy and as vaccines. The present
CC   sequence is HIV pol protein
XX
SQ   Sequence 1003 AA;

Alignment Scores:			
Pred. No.:	4.9e-199	Length:	1003
Score:	3535.00	Matches:	659
Percent Similarity:	96.12%	Conservative:	34
Best Local Similarity:	91.40%	Mismatches:	24
Query Match:	77.13%	Indels:	4
DB:	6	Gaps:	2

US-09-610-313B-30 (1-2469) x AAO30963 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||   |||||||||||||||||||   ||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
              ||||||||||||||||   ||||||||||||||      |||   :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
              ||||||||||||:::||||||||||||:::      |||||||||||:::|||
Db         41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||:::||||||:::|||||||||:::||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              |||||||||||   ||||||||||||||||||||:::|||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              |||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||||      ||||||    ||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
              ||||||||||||||||||||    :::||||||:::||||||||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
              ||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
              |||:::||||||||||||||||||   |||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
```

```
Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
                    ||||||||||::::::::||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
              :::||||||||||||||||||||      |||||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
              ||||||||||||||||||||||||||||||:::||||||   ||||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
              |||||||||||||||||||||||||||||        ||||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
              ||||||||||||||||||:::|||||||||||||||||||||:::|||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
              :::|||:::||||||:::|||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
              ||||||:::||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370
              |||
Db        721 Asp 721
```

RESULT 13
ADN36413
ID   ADN36413 standard; protein; 1003 AA.
XX
AC   ADN36413;
XX
DT   15-JUL-2004 (first entry)
XX
DE   HIV protein for anti-HIV vaccine.
XX
KW   anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO2004035006-A2.
XX
PD   29-APR-2004.
XX
PF   17-OCT-2003; 2003WO-US033112.
XX
PR   18-OCT-2002; 2002US-0419465P.
XX
PA   (AARO-) AARON DIAMOND AIDS RES CENT.
XX
PI   Huang Y, Ho DD, Chen Z;
XX
DR   WPI; 2004-348328/32.
DR   N-PSDB; ADN36412.
XX
PT   Nucleic acid vector comprising at least one HIV sequence operably linked
PT   to a promoter and encoding a protein that does not assemble into viral
PT   particles, useful in immunizing a subject against HIV infection.
XX
PS   Disclosure; SEQ ID NO 27; 166pp; English.
XX
CC   The invention relates to a nucleic acid vector comprising at least one
CC   HIV sequence operably linked to a promoter and encoding a protein that
CC   does not assemble into viral particles. The nucleic acid vector is useful
CC   in immunizing a subject against HIV infection. This sequence corresponds
CC   to a nucleic acid used in the invention.
XX
SQ   Sequence 1003 AA;

Alignment Scores:

Pred. No.:	4.9e-199	Length:	1003
Score:	3535.00	Matches:	659
Percent Similarity:	96.12%	Conservative:	34
Best Local Similarity:	91.40%	Mismatches:	24
Query Match:	77.13%	Indels:	4
DB:	8	Gaps:	2

US-09-610-313B-30 (1-2469) x ADN36413 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||   |||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                 ||||||||||||||||||   |||||||||||||||      |||   :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||:::||||||||||||:::         ||||||||||||:::|||
Db         41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||||:::||||||:::||||||||||:::||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
```

```
Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
Db    321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy   1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
Db    341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy   1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
Db    361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy   1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
Db    381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy   1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
Db    401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy   1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
Db    421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy   1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
Db    441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy   1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
Db    461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy   1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
Db    481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy   1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
Db    501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy   1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
Db    521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy   1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
Db    541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy   1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
Db    561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy   1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
Db    581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy   2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
Db    601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy   2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
Db    621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy   2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
Db    641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy   2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
Db    661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy   2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
Db    681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy   2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
Db    701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy   2368 GAT 2370
Db    721 Asp 721
```

```
RESULT 14
AAP81861
ID   AAP81861 standard; protein; 1002 AA.
XX
AC   AAP81861;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   16-DEC-1990  (first entry)
XX
DE   Sequence encoded by LAV MA L POL gene.
XX
KW   HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.
XX
OS   Lymphadenopathy-associated virus.
XX
PN   WO8707906-A.
XX
PD   30-DEC-1987.
XX
PF   22-JUN-1987;   87WO-EP000326.
XX
```

```
PR   23-JUN-1986;   86EP-00401380.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Alizon M,  Sonigo P,  Wainhobson S,  Montagnier L,  Alizon M;
PI   Sonigo P,  Wainhobson S,  Montagnier L;
XX
DR   WPI; 1988-014396/02.
DR   N-PSDB; AAN80437.
XX
PT   New variants of lymphadenopathy associated virus (LAV) - used for prodn.
PT   of DNA, antigens and antibodies used in diagnosis of AIDS and pre-AIDS.
XX
PS   Claim 8; Fig 8A-8I; 72pp; English.
XX
CC   LAV EL I (AAN80436) and LAV MA L (AAN80437) were isolated from the
CC   peripheral blood lymphocytes of patients. Different AIDS virus isolates
CC   concerned are designated by 3 letters of the patients name. Stable probes
CC   including the DNA sequences can be used for detection of the new LAV
CC   viruses or related viruses or DNA proviruses in eg. biological samples.
CC   The proteins or peptides can be used for detection of antibodies induced
CC   in vivo and present in biological fluids. The DNA can also be used for
CC   the expression of LAV viral antigens for the prodn. of a vaccine against
CC   LAV. The polypeptides can also be used for the prodn. of antibodies for
CC   the detection of proteins related to the LAV viruses, partic. for
CC   diagnosis of AIDS or pre-AIDS. (Updated on 25-MAR-2003 to correct PI
CC   field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 1002 AA;

Alignment Scores:
Pred. No.:              6.87e-199      Length:         1002
Score:                  3532.50        Matches:        656
Percent Similarity:     96.53%         Conservative:   39
Best Local Similarity:  91.11%         Mismatches:     22
Query Match:            77.08%         Indels:         3
DB:                     1              Gaps:           2

US-09-610-313B-30 (1-2469) x AAP81861 (1-1002)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
                 ||||||||||||||||||||||||||||:::|||      |||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACCCTG 390
              |||||   ||||||||||||||||       :::    :::|||||||||||||||||
Db         41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPheSerPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||:::|||:::::::::|||||||||||:::|||||||||||||||||||||
Db         61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              |||||||||||||||||||||||||||||:::::::||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||||||||||||||||||||:::|||:::||||||||||||||||||||||||
Db        121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              ||||||||||||||||||||||||||||   ||||||::::::|||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230
              |||||||||||||||||     :::|||||||||||||||||||||||||||||||||||
Db        321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy       1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290
              ||||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy       1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350
              :::||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy       1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1410
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400

Qy       1411 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1470
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy       1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440

Qy       1531 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1590
              ||||||||||||||||||||||||||||||||   ||||||||||||||||   ||||||
Db        441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuProGluAsn 460

Qy       1591 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1650
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1651 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1710
              |||||||||||||||||||      |||||||||||||||||||||||||||   :::
Db        481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500

Qy       1711 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1770
              ||||||||||||||||||||||||||||:::::::::::|||||||||||||||||||||
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520

Qy       1771 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1830
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
```

```
Db        521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540

Qy       1831 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890
              ||||||||||||||||||||||||||||||||||||      |||||||||:::||||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy       1891 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1951 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2010
              ||||||||||||   ||||||||:::||||||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       2011 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2070
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||:::
Db        601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2071 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy       2131 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2191 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2250
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

Qy       2251 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2310
              :::|||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
              |||:::||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 15
ABR55489
ID   ABR55489 standard; protein; 1003 AA.
XX
AC   ABR55489;
XX
DT   11-AUG-2003 (first entry)
XX
DE   Amino acid sequence of a HIV pol protein.
XX
KW   Polymorphic marker; host response; HIV; AIDS related virus; vaccine;
KW   hepatitis related virus; HCV; HBV; drug resistance; pol.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003035097-A1.
XX
PD   01-MAY-2003.
XX
PF   23-OCT-2002; 2002WO-AU001450.
XX
PR   23-OCT-2001; 2001AU-00008425.
XX
PA   (EPIP-) EPIPOP PTY LTD.
XX
PI   Mallal S;
XX
DR   WPI; 2003-449231/42.
XX
PT   Determining the influence of variation in host genes on the selection of
PT   microorganisms with protein substitutions, comprises typing individuals
PT   of a cohort infected with a microorganism for an intrinsic polymorphic
PT   marker.
XX
PS   Claim 22; Page 88-89; 157pp; English.
XX
CC   The specification describes a method of determining the influence of
CC   variation in host genes on selection of microorganisms with protein
CC   substitutions. The method comprises typing all individuals of a
CC   population of patients infected with a microorganism for at least one
CC   selected intrinsic polymorphic marker involved in the host response to
CC   the presence of the microorganism. The method is useful for examining
CC   selective pressures confronting a wide range of organisms that exhibit
CC   pathogenic traits in a host, such as bacteria, fungi, mycobacterium,
CC   viruses and virus-like particles; for examining microorganisms that have
CC   adapted to evolve rapidly, including HIV and AIDS related viruses and the
CC   hepatitis related viruses such as HCV and HBV. The method is useful for
CC   designing a vaccine to prevent or delay the emergence of drug resistance
CC   in patients treated with a particular drug specific for a microorganism
CC   where the drug affects the replication of the microorganism at the
CC   nucleotide or amino acid level. The present sequence represents a HIV pol
CC   protein, which is expected to provide optimal cytotoxic T lymphocyte
CC   (CTL) induced therapeutic protection to the cohort examined in that study
XX
SQ   Sequence 1003 AA;

```
Alignment Scores:
Pred. No.:              7.35e-199     Length:         1003
Score:                  3532.00       Matches:        658
Percent Similarity:     96.26%        Conservative:   36
Best Local Similarity:  91.26%        Mismatches:     23
Query Match:            77.07%        Indels:         4
DB:                     6             Gaps:           2

US-09-610-313B-30 (1-2469) x ABR55489 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::|||||||||||||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGAC------AACCCC 333
                 ||||||||||||||||   ||||||||||||||   |||:::      |||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGluAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||:::|||||||||||||:::      :::|||||||||||||
Db         41 ThrSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||||:::||||||||||||:::||||||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||:::|||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleIleIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              ||||||||||||   |||||||||||||||||||||:::|||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||         ||||||   ||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGAC 1227
              |||||||||||||||||||||   :::||||||:::||||||||||||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1228 GACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1287
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1288 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1347
              |||:::||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1348 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||   
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1408 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1467
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1468 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1527
              ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1528 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1587
                 ||||||||||||:::::::::||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1588 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1647
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1707
              :::|||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1708 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1767
              ||||||||||||||||||||||||||||||:::||||||   ||||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy       1768 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
              |||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy       1828 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
              ||||||||||||:::||||||||||||||||||||||||||||   ||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560

Qy       1888 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1948 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2007
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2008 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2067
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy       2068 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2127
              :::||||||||||||:::||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2128 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2188 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2247
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2248 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2307
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2308 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2367
              ||||||:::|||||||||||||||   |||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2368 GAT 2370
              |||
Db        721 Asp 721

Search completed: June  2, 2005, 04:41:43
Job time : 256.242 secs
```

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  1, 2005, 11:33:41 ; Search time 167.833 Seconds
                                          (without alignments)
                                          11351.628 Million cell updates/sec

Title:          US-09-610-313B-31
Perfect score:  4573
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2463

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters:      4211384

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155136_15139/app_query.fasta_1.7893
-DB=A_Geneseq_16Dec04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_557_@runat_31052005_155136_15139 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      A_Geneseq_16Dec04:*
                1:  geneseqp1980s:*
                2:  geneseqp1990s:*
                3:  geneseqp2000s:*
                4:  geneseqp2001s:*
                5:  geneseqp2002s:*
                6:  geneseqp2003as:*
                7:  geneseqp2003bs:*
                8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3766.5	82.4	1435	8	ADS19486	Ads19486 Modified
2	3744.5	81.9	3183	8	ADP84803	Adp84803 HIV-1 hyb
3	3711.5	81.2	1457	8	ADN36406	Adn36406 Human pro
4	3620	79.2	1003	3	AAB69289	Aab69289 HIV-1 non
5	3611	79.0	999	3	AAB69286	Aab69286 HIV-1 non
6	3606	78.9	1005	3	AAB69287	Aab69287 HIV-1 non
7	3570	78.1	998	6	AAE37601	Aae37601 HIV-1 sub
8	3563	77.9	3025	4	AAB86169	Aab86169 HIV-1 sub
9	3521	77.0	854	5	AAM48949	Aam48949 HIV-1 sub
10	3517.5	76.9	1000	3	AAB69282	Aab69282 HIV-1 non
11	3499.5	76.5	1002	2	AAW72993	Aaw72993 HIV isola
12	3497	76.5	1003	6	AAO30963	Aao30963 HIV pol p
13	3497	76.5	1003	8	ADN36413	Adn36413 HIV prote
14	3494.5	76.4	1002	1	AAP81861	Aap81861 Sequence
15	3494	76.4	1003	6	ABR55489	Abr55489 Amino aci
16	3492	76.4	1003	1	AAP60420	Aap60420 Sequence
17	3492	76.4	1003	1	AAP70861	Aap70861 Sequence
18	3486	76.2	1003	3	AAY70602	Aay70602 Codon opt
19	3486	76.2	1003	3	AAY70601	Aay70601 Corrected
20	3484	76.2	1003	5	AAO19387	Aao19387 Lymphaden
21	3478.5	76.1	1003	2	AAR08060	Aar08060 HIV-1 pol
22	3475	76.0	1015	1	AAP60347	Aap60347 HTLV-III
23	3475	76.0	1015	2	AAR43867	Aar43867 HTLV-III
24	3475	76.0	1015	4	AAB85993	Aab85993 Amino aci
25	3472	75.9	1003	3	AAB10047	Aab10047 HIV-1 pol
26	3472	75.9	1003	3	AAY70600	Aay70600 Wild type
27	3471	75.9	1016	2	AAR08062	Aar08062 AcNPV-HIV
28	3468	75.8	1015	2	AAR43875	Aar43875 HTLV-III
29	3464	75.7	1016	2	AAR08063	Aar08063 HIV-1 pol
30	3461	75.7	2033	2	AAR08056	Aar08056 HIV-1 pol
31	3457	75.6	1016	2	AAR08054	Aar08054 HIV-1 pol
32	3455	75.6	1003	3	AAB69284	Aab69284 HIV-1 non
33	3454	75.5	2033	2	AAR08055	Aar08055 HIV-1 pol
34	3453	75.5	1003	8	ADP20076	Adp20076 Human imm
35	3449	75.4	739	5	AAU11874	Aau11874 HIV pol p
36	3449	75.4	1003	1	AAP61508	Aap61508 Sequence
37	3449	75.4	1003	2	AAR29705	Aar29705 pol gene
38	3449	75.4	1003	3	AAY77301	Aay77301 HIV-1 (AT
39	3449	75.4	1003	5	AAE35789	Aae35789 ARV-2 (9B
40	3449	75.4	1003	6	ABU63185	Abu63185 Pol prote
41	3449	75.4	1491	1	AAP91048	Aap91048 Transcrip
42	3446	75.4	1003	3	AAB69279	Aab69279 HIV-1 non
43	3444	75.3	1003	6	ABU57552	Abu57552 AIDS asso
44	3442	75.3	1012	6	ABU63325	Abu63325 Human lym
45	3439.5	75.2	1003	2	AAR08059	Aar08059 HIV-1 pol

ALIGNMENTS

```
RESULT 1
ADS19486
ID   ADS19486 standard; protein; 1435 AA.
XX
AC   ADS19486;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Modified HIV-1 GagPol protein.
XX
KW   Gene delivery; epithelial cell; respiratory system; mammal;
KW   lentiviral packaging system; expression vector; gagpol; gene expression;
KW   lentivirus; cystic fibrosis transmembrane conductance regulator; CFTR;
KW   cystic fibrosis; CF; deltaPD; transepithelial potential difference;
KW   gene therapy; HIV-1; mutant; mutein.
XX
OS   Human immunodeficiency virus 1.
OS   Synthetic.
XX
PN   US2004037780-A1.
XX
PD   26-FEB-2004.
XX
PF   23-AUG-2002; 2002US-00226638.
XX
PR   19-NOV-2001; 2001AU-00008942.
XX
PA   (PARS/) PARSONS D.
PA   (ANSO/) ANSON D.
PA   (LIMB/) LIMBERIS M.
PA   (FULL/) FULLER M.
XX
PI   Parsons D,  Anson D,  Limberis M,  Fuller M;
XX
```

```
DR   WPI; 2004-191004/18.
DR   N-PSDB; ADS19485.
XX
PT   Delivering exogenous (e.g. CFTR) genes to epithelial cells in the
PT   respiratory tract using a penetration agent and a lentiviral packaging
PT   system, useful for treating e.g. cystic fibrosis.
XX
PS   Disclosure; Fig 16; 45pp; English.
XX
CC   The present invention relates to a method of delivering one or more
CC   exogenous genes for expression in an epithelial cell in the respiratory
CC   system of a mammal to give persistent expression of the gene in the
CC   epithelial cell, using a lentiviral packaging (expression vector) system.
CC   Also disclosed are (i) a recombinant lentiviral packaging system ,
CC   comprising a first nucleic acid molecule including a gag gene sequence
CC   encoding a Gag protein, and a second nucleic acid molecule including a
CC   gagpol gene sequence encoding a GagPol protein (the gagpol gene sequence
CC   has degenerative nucleotide changes in the frame shift sequence AUUUUUU
CC   to reduce the chance of a frame shift which switches expression of the
CC   Gagpol protein to the Gag protein in wild type lentivirus and the
CC   packaging system additionally comprises a lentiviral vector nucleic acid
CC   molecule which does not encode either the gag gene or the gagpol gene or
CC   both, and (ii) a recombinant nucleic acid molecule encoding a lentiviral
CC   gagpol gene having degenerative nucleotide changes in the frame shift
CC   sequence AUUUUUU to reduce the chance of a frame shift which switches
CC   expression of the Gagpol protein to the Gag protein in the wild type
CC   lentivirus. The exogenous gene is expressed in sufficient numbers of
CC   cells and amounts to provide an ameliorating effect for a respiratory
CC   condition. The exogenous gene is cystic fibrosis transmembrane
CC   conductance regulator (CFTR) and the condition is cystic fibrosis
CC   ((CF)the expression of the CFTR gene is sufficient to provide a
CC   significant shift of a reduced deltaPD (transepithelial potential
CC   difference) back to normal levels in the mammal). The cell is non-
CC   terminally differentiated and is capable of differentiating into 2 or
CC   more cell classes e.g. ciliated cells, non-ciliated cells, secretory
CC   cells and basal cells. The exogenous gene may be enzymic. The gagpol gene
CC   sequence has additionally degenerate nucleotide substitutions which
CC   destabilise the hairpin structure associated with the frame shift event.
CC   The gag gene is a truncation of the wild type gagpol gene so that it can
CC   no longer be translated to form Gagpol. The lentivirus is HIV or HIV
CC   derived. The method and recombinant lentiviral packaging system may be
CC   used for delivering nucleic acids to epithelial cells in the respiratory
CC   system of a mammal. In particular they may be used to deliver the CFTR
CC   gene for the treatment of CF. It has been found that the delivery of a
CC   recombinant lentivirus carrying an exogenous gene to the respiratory
CC   system, following the delivery of a non-toxic amount of a penetration
CC   enhancer can provide persistence of expression of a gene product. The
CC   present sequence represents modified HIV-1 GagPol protein.
XX
SQ   Sequence 1435 AA;

Alignment Scores:
Pred. No.:              2.87e-209     Length:        1435
Score:                  3766.50       Matches:       702
Percent Similarity:     94.31%        Conservative:  44
Best Local Similarity:  88.75%        Mismatches:    38
Query Match:            82.36%        Indels:        7
DB:                     8             Gaps:          4

US-09-610-313B-31 (1-2463) x ADS19486 (1-1435)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACC---AGCGCCAACATCCTGATGCAGCGCAGCAAC 69
                 :::|||||||||||||||||    |||   ||||||    |||:::||||||||    |||
Db        363 LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn 382

Qy         70 TTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCC 129
              |||:::      ::::::   :::|||||||||||||||||||||||||||||||||
Db        383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla 402

Qy        130 CGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAG 189
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422

Qy        190 ATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAG 249
              ||||||||||||||||||||||||||||||||||      |||||||||||||||||||||||
Db        423 MetLysAspCysThrGluArgGlnAlaAsnPheLeuArgGluAspLeuAlaPheProGln 442

Qy        250 GGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAG 309
              ||||||||||||:::|||    |||||||||    ||||||||||||||||      ||||||
Db        443 GlyLysAlaArgLysPheSerSerGluGlnThrArgAlaAsnSerProIleArgArgGlu 462

Qy        310 CTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGC 363
                 ||||||          |||    :::|||      ||||||||||||||:::||||||||
Db        463 ArgGlnValTrpArgArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGly 482

Qy        364 ACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAG 417
              |||:::          :::||||||||||||||||||||||||||||||||||:::||||||
Db        483 ThrValSerPheSerPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLys 502

Qy        418 GTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAG 477
              :::|||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        503 IleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGlu 522

Qy        478 GAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATC 537
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        523 GluMetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIle 542

Qy        538 AAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACC 597
              ||||||||||||||||||||||||||    |||||||||||||||    |||||||||||||||
Db        543 LysValArgGlnTyrAspGlnIleProIleGluIleCysGlyHisLysAlaIleGlyThr 562

Qy        598 GTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGC 657
              ||||||:::|||||||||||||||||||||||||||||||||:::|||||||||:::||||
Db        563 ValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGly 582

Qy        658 TGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGC 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 CysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGly 602

Qy        718 ATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACC 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 MetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuVal 622

Qy        778 GCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCC 837
                 ||||||      |||||||||||||||||||||:::||||||||||||||||||||||
Db        623 GluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnPro 642

Qy        838 TACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTG 897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 TyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuVal 662

Qy        898 GACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC 957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 AspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIlePro 682

Qy        958 CACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTAC 1017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 HisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyr 702

Qy       1018 TTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATC 1077
              |||||||||||||||    |||||||||||||||||||||||||||||||||||||||||
Db        703 PheSerValProLeuHisGluAspPheArgLysTyrThrAlaPheThrIleProSerIle 722

Qy       1078 AACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGC 1137
              ||||||||||||||||    ||||||||||||||||||||||||||||||||||||||||
Db        723 AsnAsnGluThrProGlyThrArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGly 742

Qy       1138 AGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAAC 1197
              ||||||:::|||||||||||||||||||||    ||||||||||||||||||    :::|||
Db        743 SerProAlaIlePheGlnSerSerMetThrThrIleLeuGluProPheArgLysGlnAsn 762

Qy       1198 CCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATC 1251
              |||::::::||||||||||||          ||||||||||||||||||||||||||||
Db        763 ProAspLeuValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIle 782

Qy       1252 GGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACC 1311
```

```
               |||||||||||   ||||||||||||||||:::||||||||||||||||||||||
Db         783 GlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThr 802

Qy        1312 ACCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCCTGTGGATGGGCTACGAGCTGCAC 1371
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         803 ThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHis 822

Qy        1372 CCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAAC 1431
               |||||||||||||||||||||||||||   |||||||||:::||||||||||||||
Db         823 ProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsn 842

Qy        1432 GACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAG 1491
               ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||
Db         843 AspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys 862

Qy        1492 GTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTG 1551
               ||||||||||||||||||||||||||||||   |||||||||||::::::::||||||
Db         863 ValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeu 882

Qy        1552 ACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCAC 1611
               ||||||||||||||||||||||||||||||||||||||||||||:::|||||||||||
Db         883 ThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHis 902

Qy        1612 GGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGAC 1671
               |||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db         903 GlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGly 922

Qy        1672 CAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCC 1731
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         923 GlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAla 942

Qy        1732 AAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATC 1791
               :::   |||   |||||||||||||||||||||||||||||||||||||||||||||||
Db         943 ArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIle 962

Qy        1792 GCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAG 1851
               |||   ||||||||||||||||||||||||||||||||:::|||||||||||||||
Db         963 AlaThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLys 982

Qy        1852 GAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAG 1911
               |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db         983 GluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGlu 1002

Qy        1912 TTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATC 1971
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1003 PheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIle 1022

Qy        1972 GGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCC 2031
               ||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        1023 GlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAla 1042

Qy        2032 GGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAG 2091
               |||||||||||::::::||||||||||||:::||||||||||||:::|||||||||||
Db        1043 GlyTyrValThrAsnLysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGln 1062

Qy        2092 AAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATC 2151
               ||||||||||||||||||||   ||||||||||||||||||||||   ||||||||||
Db        1063 LysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIle 1082

Qy        2152 GTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGC 2211
               ||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        1083 ValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspArgSerGluSer 1102

Qy        2212 GAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGG 2271
               ||||||||:::||||||||||||||||||||||||||||||||||||||||:::|||
Db        1103 GluLeuValSerGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrp 1122

Qy        2272 GTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGC 2331
               ||||||||||||||||||||||||||||||||||:::||||||||||||||    |||
Db        1123 ValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGly 1142

Qy        2332 ATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
               |||||||||||||||||||||||||||||||||
Db        1143 IleArgLysValLeuPheLeuAspGlyIleAsp 1153

RESULT 2
ADP84803
ID   ADP84803 standard; protein; 3183 AA.
XX
AC   ADP84803;
XX
DT   23-SEP-2004  (first entry)
XX
DE   HIV-1 hybrid isolate pNL4-3 protein.
XX
KW   eliciting; inducing; immune response; HIV; antigen; non-pathogenic;
KW   vaccination; vaccine; HIV-I.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO2004056391-A1.
XX
PD   08-JUL-2004.
XX
PF   19-DEC-2003; 2003WO-AU001705.
XX
PR   20-DEC-2002; 2002AU-00953556.
PR   17-SEP-2003; 2003AU-00905067.
XX
PA   (UYNE-) UNIV NEW SOUTH WALES.
XX
PI   Kent SJ,  Purcell DF,  Boyle DB,  Ramsay A,  Thomson S,  Ramshaw IA;
XX
DR   WPI; 2004-500267/47.
DR   N-PSDB; ADP84804.
XX
PT   Eliciting or inducing in a mammal an immune response against HIV-I
PT   subtype AE, B or C by administering sequential doses of a recombinant
PT   plasmid and viral vectors containing the nucleic acid molecules encoding
PT   the HIV antigens.
XX
PS   Disclosure; Fig 12; 280pp; English.
XX
CC   The invention relates to a novel method for eliciting or inducing in a
CC   mammal an immune response directed to a virus, preferably Human
CC   immunodeficiency virus (HIV). The method comprises sequentially
CC   administering to the mammal one or more sequential doses of a recombinant
CC   plasmid vector or recombinant viral vector or its derivative, into which
CC   the nucleic acid molecules encoding all, part or a modified form of two
CC   or more antigens of the virus are incorporated, and one or more optimized
CC   CpG motifs, where the antigens have been rendered substantially non-
CC   pathogenic. The invention further comprises: a method for treating or
CC   preventing viral infection in a mammal; a method of vaccinating a mammal
CC   against a viral pathogen; a method of eliciting or inducing, in a mammal,
CC   an immune response directed to HIV; a vaccine capable of inducing an
CC   immune response directed to a virus comprising the recombinant plasmid
CC   vector or recombinant viral vector or its functional derivative; a
CC   nucleic acid construct or its functional derivative comprising the
CC   plasmid vector or recombinant viral vector; a plasmid vector; a
CC   recombinant viral vector; a pharmaceutical composition comprising the
CC   nucleic acid, plasmid vector or recombinant viral vector constructs. The
CC   method is useful in eliciting or inducing in a mammal an immune response
CC   against HIV-I subtype AE, B or C. This sequence represents the protein of
CC   the HIV-1 hybrid isolate pNL4-3 DNA of the invention.
XX
SQ   Sequence 3183 AA;

Alignment Scores:
Pred. No.:              6.39e-208     Length:        3183
Score:                  3744.50       Matches:       707
Percent Similarity:     87.43%        Conservative:  44
Best Local Similarity:  82.31%        Mismatches:    33
Query Match:            81.88%        Indels:        75
DB:                     8             Gaps:          6

US-09-610-313B-31 (1-2463) x ADP84803 (1-3183)
```

```
Qy     13 ATGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAAC 69
          :::||||||||||||||||||      |||:::   |||   |||::::::|||:::   |||
Db    363 LeuAlaGluAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLysGlyAsn 382

Qy     70 TTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCC 129
          |||:::      ::::::   :::|||||||||||||||||||||||||||||||||||||
Db    383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla 402

Qy    130 CGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAG 189
          :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422

Qy    190 ATGAAGGACTGCACCGAGCGCCAGGCCAAC------------------------------ 219
          ||||||||||||||||||||||||||||||
Db    423 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys 442

Qy    220 ---------------------------------------------TTCTTCCGCGAGGAC 234
                                                       |||||||||||||||
Db    443 GlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPhePheArgGluAsp 462

Qy    235 CTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGC 294
          ||||||||||||||||||||||||||||||||||||   |||||||||   ||||||||||||
Db    463 LeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSer 482

Qy    294 ------------------------------------------------------------ 294

Db    483 ProProGluGluSerPheArgPheGlyGluGluThrThrThrProSerGlnLysGlnGlu 502

Qy    295 ------------------------------CCCACCAGCCGCGAGCTGCAGGTG------ 318
                                        ||||||   |||||||||||||||
Db    503 ProIleAspLysGluLeuTyrProLeuAlaProThrArgArgGluLeuGlnValTrpGly 522

Qy    319 CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC------ 372
          |||   :::|||         |||||||||||||||:::||||||||||||::::::
Db    523 ArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSer 542

Qy    373 ---------------------------------------TTCCCCCAGATCACCCTGTGG 393
                                                 |||||||||||||||||||||
Db    543 SerLeuArgSerLeuPheGlySerAspProSerSerGlnPheProGlnIleThrLeuTrp 562

Qy    394 CAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACC 453
          |||||||||||||||:::||||||:::|||||||||:::||||||||||||||||||||||
Db    563 GlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThr 582

Qy    454 GGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATG 513
          |||||||||||||||||||||||||||||:::|||||||||:::||||||||||||||||
Db    583 GlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysProLysMet 602

Qy    514 ATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATC 573
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 IleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIle 622

Qy    574 TGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGC 633
          ||||||   |||||||||||||||||||||||:::||||||||||||||||||||||||||
Db    623 CysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGly 642

Qy    634 CGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACC 693
          ||||||:::|||||||||:::||||||||||||||||||||||||||||||||||||||||
Db    643 ArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThr 662

Qy    694 GTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACC 753
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    663 ValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThr 682

Qy    754 GAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATC 813
          ||||||||||||||||||||||      ||||||   |||||||||||||||||||||||
Db    683 GluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIle 702

Qy    814 ACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGAC 873
          :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    703 SerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAsp 722

Qy    874 AGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTC 933
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    723 SerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPhe 742

Qy    934 TGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACC 993
          ||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db    743 TrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSerValThr 762

Qy    994 GTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTAC 1053
          |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db    763 ValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyr 782

Qy   1054 ACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAAC 1113
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    783 ThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsn 802

Qy   1114 GTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATC 1173
          ||||||||||||||||||||||||||:::|||||||||   ||||||||||||||||
Db    803 ValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThrLysIle 822

Qy   1174 CTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTG 1227
          ||||||||||||||   :::||||||:::||||||||||||||||          |||
Db    823 LeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeu 842

Qy   1228 TACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAG 1287
          ||||||||||||||||||||||||||||||||||   ||||||||||||||||||:::
Db    843 TyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGln 862

Qy   1288 CACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTC 1347
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    863 HisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPhe 882

Qy   1348 CTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCC 1407
          ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db    883 LeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleValLeuPro 902

Qy   1408 GAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCC 1467
          ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    903 GluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAla 922

Qy   1468 AGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAG 1527
          |||||||||||||   ||||||||||||||||||||||||||||||||||||||    |||
Db    923 SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLys 942

Qy   1528 GCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGC 1587
          |||||||||::::::||||||||||||||||||||||||||||||||||||||||||||||
Db    943 AlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArg 962

Qy   1588 GAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCC 1647
          |||||||||:::|||||||||||||||||||||||||||||||||||||||:::||||||
Db    963 GluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAla 982

Qy   1648 GAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAG 1707
          ||||||||||||||||||         |||||||||||||||||||||||||||||||||
Db    983 GluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLys 1002

Qy   1708 AACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAG 1767
          |||||||||||||||||||||||||:::|||:::   |||||||||||||||||||||||
Db   1003 AsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspValLysGln 1022

Qy   1768 CTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCC 1827
          |||||||||||||||||||||||||||     |||||||||||||||||||||||||||||
Db   1023 LeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThrPro 1042

Qy   1828 AAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAG 1887
          ||||||:::||||||||||||||||||||||||||     |||||||||:::|||||||||
Db   1043 LysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrpGln 1062

Qy   1888 GCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTAC 1947
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1063 AlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyr 1082

Qy   1948 CAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAAC 2007
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1083 GlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAlaAsn 1102

Qy      2008 CGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTG 2067
             ||||||||||||:::||||||||||||||||||||||||||||||||||||||:::|||
Db      1103 ArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValVal 1122

Qy      2068 AGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAG 2127
                ||||||:::|||||||||||||||||||||||||||||||||   ||||||||||||
Db      1123 ProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGln 1142

Qy      2128 GACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAG 2187
             |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1143 AspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGln 1162

Qy      2188 GCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAG 2247
             ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db      1163 AlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleLys 1182

Qy      2248 AAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAG 2307
             |||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db      1183 LysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGln 1202

Qy      2308 ATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
             :::|||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db      1203 ValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1221
```

```
RESULT 3
ADN36406
ID   ADN36406 standard; protein; 1457 AA.
XX
AC   ADN36406;
XX
DT   15-JUL-2004  (first entry)
XX
DE   Human protein for anti-HIV vaccine.
XX
KW   anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX
OS   Homo sapiens.
XX
PN   WO2004035006-A2.
XX
PD   29-APR-2004.
XX
PF   17-OCT-2003; 2003WO-US033112.
XX
PR   18-OCT-2002; 2002US-0419465P.
XX
PA   (AARO-) AARON DIAMOND AIDS RES CENT.
XX
PI   Huang Y,  Ho DD,  Chen Z;
XX
DR   WPI; 2004-348328/32.
DR   N-PSDB; ADN36405.
XX
PT   Nucleic acid vector comprising at least one HIV sequence operably linked
PT   to a promoter and encoding a protein that does not assemble into viral
PT   particles, useful in immunizing a subject against HIV infection.
XX
PS   Disclosure; SEQ ID NO 20; 166pp; English.
XX
CC   The invention relates to a nucleic acid vector comprising at least one
CC   HIV sequence operably linked to a promoter and encoding a protein that
CC   does not assemble into viral particles. The nucleic acid vector is useful
CC   in immunizing a subject against HIV infection. This sequence corresponds
CC   to a peptide used in the invention.
XX
SQ   Sequence 1457 AA;

Alignment Scores:
Pred. No.:            4.34e-206     Length:       1457
Score:                3711.50       Matches:      706
Percent Similarity:     91.86%      Conservative:    28
Best Local Similarity:  88.36%      Mismatches:      45
Query Match:            81.16%      Indels:          21
DB:                     8           Gaps:            4

US-09-610-313B-31 (1-2463) x ADN36406 (1-1457)

Qy        13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
             :::||||||||||||||||||      :::      ||||||||||||||||||||||||
Db       383 LeuAlaGluAlaMetSerGlnAla---AsnGlyThrIleLeuMetGlnArgSerAsnPhe 401

Qy        73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
             ||||||   ||||||||:::||||||||||||||||||||||||||||||||||||||||
Db       402 LysGlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 421

Qy       133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       422 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 441

Qy       193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
             |||||||||||||||||||||||||||||||||     :::    ||||||    :::
Db       442 LysAspCysThrGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProProHisLysGl 461

Qy       253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
                ||||||  :::       ||||||   ||||||||||||||||||     |||
Db       461 yArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPh 481

Qy       313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
                         |||||||||||   :::      |||     |||   |||    ||
Db       481 eGlyPheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuTh 501

Qy       373 TTC------------------------------------CCCCAGATCACCCTGTGGCA 395
             |                                      ||||||||||||||||||
Db       501 rSerLeuLysSerLeuPheGlySerAspProLeuSerGlnProGlnIleThrLeuTrpGl 521

Qy       396 GCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGG 455
             ||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db       521 nArgProLeuValSerIleArgValGlyGlyGlnIleLysGluAlaLeuLeu-------- 538

Qy       456 CGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGAT 515
                 |||||||||||||||||||:::::||||||||||||||||||||||||||||||||
Db       539 ----AspAspThrValLeuGluGluValAsnLeuProGlyLysTrpLysProLysMetIl 557

Qy       516 CGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTG 575
             ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db       557 eGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIleGluIleCy 577

Qy       576 CGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCG 635
             |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db       577 sGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyAr 597

Qy       636 CAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGT 695
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       597 gAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrIl 617

Qy       696 GCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGA 755
             :||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db       617 eProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrGl 637

Qy       756 GGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCAC 815
             ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db       637 uGluLysIleLysAlaLeuThrAlaIleCysAspGluMetGluLysGluGlyLysIleTh 657

Qy       816 CAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAG 875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       657 rLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSe 677

Qy       876 CACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTG 935
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       677 rThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTr 697

Qy       936 GGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGT 995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        697 pGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVa 717

Qy        996 GCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACAC 1055
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        717 lLeuAspValGlyAspAlaTyrPheSerValProLeuTyrGluAspPheArgLysTyrTh 737

Qy       1056 CGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGT 1115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 rAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVa 757

Qy       1116 GCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175
              ||||||||||||||||||||||||||:::||||||||||   ||||||   ||||||||
Db        757 lLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetAlaLysIleLe 777

Qy       1176 GGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTA 1229
              ||||||||||||||:::|||||||||||||||||||||||||            |||||
Db        777 uGluProPheArgAlaGlnAsnProGluIleValIleTyrGlnTyrGlyAspAspLeuTy 797

Qy       1230 CGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCA 1289
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:::||
Db        797 rValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHi 817

Qy       1290 CCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCT 1349
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 sLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLe 837

Qy       1350 GTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGA 1409
              |||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db        837 uTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGl 857

Qy       1410 GAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAG 1469
              ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 uLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSe 877

Qy       1470 CCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGC 1529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 rGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAl 897

Qy       1530 CCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGA 1589
              |||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        897 aLeuThrAspIleIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGl 917

Qy       1590 GATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGA 1649
              ||||||:::|||||||||||||||   ||||||||||||||||||||||:::||||||
Db        917 uIleLeuLysGluProValHisGlyAlaTyrTyrAspProSerLysAspLeuIleAlaGl 937

Qy       1650 GATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAA 1709
              ||||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db        937 uIleGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAs 957

Qy       1710 CCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCT 1769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 nLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLe 977

Qy       1770 GACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAA 1829
              ||||||||||||||||||||:::||||||||||||||||||||||||||   |||||
Db        977 uThrGluAlaValGlnLysIleSerMetGluSerIleValIleTrpGlyLysIleProLy 997

Qy       1830 GTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGC 1889
              |||||||||||||||   |||||||||||||||||   ||||||   ||||||||||
Db        997 sPheArgLeuProIleProLysGluThrTrpGluThrArgTrpThrAlaTyrTrpGlnAl 1017

Qy       1890 CACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCA 1949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 aThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGl 1037

Qy       1950 GCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCG 2009
              |||||||||:::||||||   |||   |||||||||||||||||||||||||||||||
Db       1037 nLeuGluLysAspProIleAlaGlyValGluThrPheTyrValAspGlyAlaAlaAsnAr 1057

Qy       2010 CGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAG 2069
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db       1057 gGluThrLysMetGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSe 1077

Qy       2070 CCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGA 2129
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db       1077 rLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleCysLeuAlaLeuGlnAs 1097

Qy       2130 CAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGC 2189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1097 pSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAl 1117

Qy       2190 CCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAA 2249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1117 aGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLy 1137

Qy       2250 GGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGAT 2309
              ||||:::||||||||||||||||||||||||||||||||||||||||||||||||||::
Db       1137 sGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVa 1157

Qy       2310 CGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              :|||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db       1157 lAspLysLeuValSerAsnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1175

RESULT 4
AAB69289
ID   AAB69289 standard; protein; 1003 AA.
XX
AC   AAB69289;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 94IN476-104 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1003 AA;

Alignment Scores:
Pred. No.:              7.76e-201      Length:        1003
Score:                  3620.00        Matches:       677
Percent Similarity:     98.05%         Conservative:  26
Best Local Similarity:  94.42%         Mismatches:    12
```

```
Query Match:          79.16%     Indels:      2
DB:                   3          Gaps:        1

US-09-610-313B-31 (1-2463) x AAB69289 (1-1003)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db      1 PhePheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerLysGln 20

Qy    280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
Db     21 AlaArgAlaAsnSerProThrSerArgGluLeuGlnValGlnGlyAspAsnProArgSer 40

Qy    340 GAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGC 399
Db     41 GluAlaGlyValGluArgGlnGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg 60

Qy    400 CCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCC 459
Db     61 ProLeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla 80

Qy    460 GACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGC 519
Db     81 AspAspThrValLeuGluGluIleAlaLeuProGlyArgTrpLysProLysMetIleGly 100

Qy    520 GGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGC 579
Db    101 GlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGly 120

Qy    580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
Db    121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsp 140

Qy    640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
Db    141 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 160

Qy    700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
Db    161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180

Qy    760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
Db    181 LysIleLysAlaLeuThrGluIleCysLysGluMetGluLysGluGlyLysIleThrLys 200

Qy    820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
Db    201.IleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysArgLysAspSerThr 220

Qy    880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
Db    221 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 240

Qy    940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
Db    241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 260

Qy   1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
Db    261 AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheGlyLysTyrThrAla 280

Qy   1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
Db    281 PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 300

Qy   1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
Db    301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320

Qy   1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTG 1233
Db    321 ProPheArgAlaArgAsnProLysIleValIleTyrGlnTyrMetAspAspLeuTyrVal 340

Qy   1234 GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTG 1293
Db    341 GlySerAspLeuGluIleGlyHisHisArgAlaLysIleGluGluLeuArgAlaHisLeu 360

Qy   1294 CTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGG 1353
Db    361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 380

Qy   1354 ATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAG 1413
Db    381 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleLysLeuProGluLys 400

Qy   1414 GAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG 1473
Db    401 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln 420

Qy   1474 ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTG 1533
Db    421 IleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440

Qy   1534 ACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATC 1593
Db    441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIle 460

Qy   1594 CTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATC 1653
Db    461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480

Qy   1654 CAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTG 1713
Db    481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500

Qy   1714 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACC 1773
Db    501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520

Qy   1774 GAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTC 1833
Db    521 GluAlaValGlnLysIleAlaIleGluSerIleValIleTrp***LysThrProLysPhe 540

Qy   1834 CGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACC 1893
Db    541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThr 560

Qy   1894 TGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTG 1953
Db    561 TrpIleProAspTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeu 580

Qy   1954 GAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAG 2013
Db    581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600

Qy   2014 ACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTG 2073
Db    601 ThrLysValGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSerLeu 620

Qy   2074 ACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGC 2133
Db    621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 640

Qy   2134 GGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAG 2193
Db    641 GlyThrGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln 660

Qy   2194 CCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAG 2253
Db    661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleAsnLysGlu 680

Qy   2254 AAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGAC 2313
Db    681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAsp 700

Qy   2314 AAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
```

```
                :::|||||||||   ||||||||||||||||||||||||||||||||||||
Db        701 ArgLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

RESULT 5
AAB69286
ID   AAB69286 standard; protein; 999 AA.
XX
AC   AAB69286;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 96ZM651-8 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 999 AA;

Alignment Scores:
Pred. No.:              2.57e-200      Length:         999
Score:                  3611.00        Matches:        677
Percent Similarity:     97.77%         Conservative:   24
Best Local Similarity:  94.42%         Mismatches:     14
Query Match:            78.96%         Indels:         2
DB:                     3              Gaps:           1

US-09-610-313B-31 (1-2463) x AAB69286 (1-999)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AlaArgAlaAsnSerProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSer 40

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGC 399
              |||||||||   ||||||||||||:::|||||||||||||||||||||||||||||||||
Db         41 GluAlaGlyValGluArgGlnGlySerLeuAsnPheProGlnIleThrLeuTrpGlnArg 60

Qy        400 CCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCC 459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ProLeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla 80

Qy        460 GACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGC 519
                 ||||||||||||||||::::::|||||||||||||||||||||||||||||||||
Db         81 GlyAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGly 100

Qy        520 GGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGC 579
              ||||||||||||||||||:::|||||||||||||||||||||     :::||||||||||
Db        101 GlyIleGlyGlyPheIleGluValArgGlnTyrAspGlnIleProMetGluIleCysGly 120

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 140

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 160

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LysIleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLys 200

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThr 220

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 240

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 260

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
              ||||||||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db        261 AspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThrAla 280

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
              |||||||||||||||||||||::::|||||||||||||||||||||||||||||||||||
Db        281 PheThrIleProSerThrAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 300

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTG 1233
              |||||||||||:::||||||:::|||||||||||||||         |||||||||
Db        321 ProPheArgAlaGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrVal 340

Qy       1234 GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTG 1293
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        341 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeu 360

Qy       1294 CTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGG 1353
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 380

Qy       1354 ATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAG 1413
              |||||||||||||||||||||||||||||||||||||||||:::|||   ||||||||
Db        381 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuAlaGluLys 400

Qy       1414 GAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG 1473
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln 420

Qy       1474 ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTG 1533
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 IleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440

Qy       1534 ACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATC 1593
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnLysGluIle 460

Qy       1594 CTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATC 1653
              |||:::|||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480

Qy       1654 CAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTG 1713
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500

Qy       1714 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACC 1773
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520

Qy       1774 GAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTC 1833
              |||||||||||||||||||||:::|||||||||||||||||||||||    |||||||||
Db        521 GluAlaValGlnLysIleAlaLeuGluSerIleValIleTrpGlyLysIleProLysPhe 540

Qy       1834 CGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACC 1893
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThr 560

Qy       1894 TGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTG 1953
              |||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        561 TrpIleProGluTrpGluPheValAsnThrProLeuLeuValLysLeuTrpTyrGlnLeu 580

Qy       1954 GAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAG 2013
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600

Qy       2014 ACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTG 2073
              ||||||:::|||||||||||||||:::|||||||||||||||||||||||||||:::|||
Db        601 ThrLysLeuGlyLysAlaGlyTyrIleThrAspArgGlyArgGlnLysIleValThrLeu 620

Qy       2074 ACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGC 2133
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSer 640

Qy       2134 GGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAG 2193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaHis 660

Qy       2194 CCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAG 2253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlu 680

Qy       2254 AAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGAC 2313
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAsp 700

Qy       2314 AAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 LysLeuValSerLysGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

RESULT 6
AAB69287
ID   AAB69287 standard; protein; 1005 AA.
XX
AC   AAB69287;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 96ZM751-3 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1005 AA;

Alignment Scores:
Pred. No.:              5e-200        Length:        1005
Score:                  3606.00       Matches:       681
Percent Similarity:     96.68%        Conservative:  18
Best Local Similarity:  94.19%        Mismatches:    16
Query Match:            78.85%        Indels:        8
DB:                     3             Gaps:          2

US-09-610-313B-31 (1-2463) x AAB69287 (1-1005)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::||||||||||||:::|||:::|||   |||   ||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGluGlyGluAlaGlyGluLeuProSerGluGln 20

Qy        280 AACCGCGCC------------------AACAGCCCCACCAGCCGCGAGCTGCAGGTGCGC 321
                  |||||                  |||||||||||||||||||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrSerSerAsnSerProThrSerArgGluLeuGlnValArg 40

Qy        322 GGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAG 381
              ||||||||||||      |||||||||||||||||||||||||||||||||   ||||||
Db         41 GlyAspAsnProCysProGluAlaGlyAlaGluArgGlnGlyThrLeuAsnCysProGln 60

Qy        382 ATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCC 441
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIleLysGluAla 80

Qy        442 CTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGG 501
              |||||||||||||||||||||||||||||||||||||||||||||::::::|||||||||
Db         81 LeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrp 100

Qy        502 AAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATC 561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIle 120

Qy        562 CTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTG 621
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        121 LeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVal 140

Qy        622 AACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGC 681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSer 160
```

```
Qy      682 CCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAG 741
            ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db      161 ProIleGluThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGln 180

Qy      742 TGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAG 801
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 TrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLys 200

Qy      802 GAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATC 861
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      201 GluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIle 220

Qy      862 AAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGC 921
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      221 LysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArg 240

Qy      922 ACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAG 981
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 ThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLys 260

Qy      982 AAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGAC 1041
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      261 LysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGly 280

Qy     1042 TTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGC 1101
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      281 PheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArg 300

Qy     1102 TACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGC 1161
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 TyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSer 320

Qy     1162 ATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG--- 1218
            |||    ||||||||||||||||||||    :::||||||||||||||||||||||
Db      321 MetIleLysIleLeuGluProPheArgThrGlnAsnProGluIleValIleTyrGlnTyr 340

Qy     1219 ---GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAG 1275
                     ||||||||||||||||||||||||||||||||||||||||||||||||
Db      341 MetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGlu 360

Qy     1276 GAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAG 1335
            |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 GluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLys 380

Qy     1336 GAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCC 1395
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      381 GluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnPro 400

Qy     1396 ATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAG 1455
            |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      401 IleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLys 420

Qy     1456 CTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTG 1515
            |||||||||   |||||||||||   ||||||||||||||||||||||||||||||||||
Db      421 LeuAsnTrp***SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeu 440

Qy     1516 CGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTG 1575
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      441 ArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeu 460

Qy     1576 GCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAG 1635
            ||||||:::||||||||||||:::||||||||||||||||||||||||||||||||||||
Db      461 AlaGluSerArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLys 480

Qy     1636 GACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAG 1695
            ||||||:::||||||||||||||||||||||||||||||||||||||:::||||||||||
Db      481 AspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnValTyrGln 500

Qy     1696 GAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAAC 1755
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      501 GluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsn 520

Qy     1756 GACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGG 1815
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      521 AspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrp 540

Qy     1816 GGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACC 1875
            ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 GlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThr 560

Qy     1876 GACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTG 1935
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      561 AspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVal 580

Qy     1936 AAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGAC 1995
            ||||||||||||||||||||||||||||||    |||||||||||:::||||||||||||
Db      581 LysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrTyrTyrValAsp 600

Qy     1996 GGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGG 2055
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArg 620

Qy     2056 CAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAG 2115
            ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db      621 GlnLysIleValThrLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGln 640

Qy     2116 CTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTG 2175
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      641 LeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeu 660

Qy     2176 GGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAG 2235
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 GlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGlu 680

Qy     2236 CAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGC 2295
            ||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db      681 GlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGly 700

Qy     2296 GGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGAC 2355
            |||||||||||:::|||||||||||||||||   ||||||||||||||||||||||||||
Db      701 GlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAsp 720

Qy     2356 GGCATCGAT 2364
            |||||||||
Db      721 GlyIleAsp 723
```

```
RESULT 7
AAE37601
ID   AAE37601 standard; protein; 998 AA.
XX
AC   AAE37601;
XX
DT   23-OCT-2003  (revised)
DT   27-AUG-2003  (first entry)
XX
DE   HIV-1 subtype C isolate Du151 reverse transcriptase (RT) protein.
XX
KW   Regulatory gene; accessory gene; HIV; human immunodeficiency virus;
KW   vaccine; infection; gene therapy; reverse transcriptase; RT; enzyme.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO2003037919-A2.
XX
PD   08-MAY-2003.
XX
PF   31-OCT-2002; 2002WO-IB004550.
XX
PR   31-OCT-2001; 2001ZA-00008978.
XX
PA   (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.
PA   (UYCA-) UNIV CAPE TOWN.
XX
PI   Williamson C,  Van Harmelen JH,  Gray CM,  Bourn W,  Karim SA;
XX
```

```
DR   WPI; 2003-430497/40.
DR   N-PSDB; AAD29258.
XX
PT   New molecules comprising HIV-1 subtype isolate regulatory/accessory
PT   genes, useful for manufacturing a vaccine for treating or preventing HIV
PT   infection.
XX
PS   Disclosure; Page 94-97; 97pp; English.
XX
CC   The invention relates to molecules comprising HIV-1 subtype isolate
CC   regulatory/accessory genes (tat, nef and rev genes) and modifications and
CC   derivatives thereof. The invention also provides proteins encoded by such
CC   genes. Sequences of the invention are useful for manufacturing vaccines
CC   for treating or preventing human immunodeficiency virus (HIV) infections.
CC   They are also useful in gene therapy. The present sequence is HIV-1
CC   subtype C isolate Du151 reverse transcriptase (RT) protein. (Updated on
CC   23-OCT-2003 to standardise OS field)
XX
SQ   Sequence 998 AA;

Alignment Scores:
Pred. No.:              6.01e-198     Length:         998
Score:                  3570.00       Matches:        666
Percent Similarity:     97.07%        Conservative:   29
Best Local Similarity:  93.02%        Mismatches:     19
Query Match:            78.07%        Indels:         2
DB:                     6             Gaps:           1

US-09-610-313B-31 (1-2463) x AAE37601 (1-998)

Qy    223 TTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAAC 282
Db      1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20

Qy    283 CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAG 342
Db     21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu 40

Qy    343 GCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
Db     41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro 60

Qy    403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
Db     61 LeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp 80

Qy    463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
Db     81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 100

Qy    523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
Db    101 IleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLys 120

Qy    583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
Db    121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140

Qy    643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
Db    141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160

Qy    703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
Db    161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180

Qy    763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
Db    181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200

Qy    823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
Db    201 GlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220

Qy    883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
Db    221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240

Qy    943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
Db    241 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp 260

Qy   1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
Db    261 ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 280

Qy   1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
Db    281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 300

Qy   1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
Db    301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320

Qy   1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGC 1236
Db    321 PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 340

Qy   1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
Db    341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeu 360

Qy   1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATG 1356
Db    361 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380

Qy   1357 GGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1416
Db    381 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400

Qy   1417 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1476
Db    401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420

Qy   1477 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1536
Db    421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440

Qy   1537 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1596
Db    441 AspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu 460

Qy   1597 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1656
Db    461 LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln 480

Qy   1657 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1716
Db    481 LysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 500

Qy   1717 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1776
Db    501 ThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThrGlu 520

Qy   1777 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1836
Db    521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540

Qy   1837 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1896
Db    541 LeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThrTrp 560

Qy   1897 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1956
Db    561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 580
```

```
Qy       1957 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2016
              |||||||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 600

Qy       2017 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2076
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||:::
Db        601 LysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeuSer 620

Qy       2077 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2136
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640

Qy       2137 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2196
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        641 SerGluValAsnIleIleThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 660

Qy       2197 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2256
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        661 AspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluArg 680

Qy       2257 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2316
              ||||||||||||||||||||||||||||||   |||:::   ||||||||    :::||||||
Db        681 ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700

Qy       2317 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              |||||||||:::||||||||||||||||||   |||||||||||||||
Db        701 LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716
```

RESULT 8

```
AAB86169
ID   AAB86169 standard; protein; 3025 AA.
XX
AC   AAB86169;
XX
DT   11-SEP-2003  (revised)
DT   21-AUG-2001  (first entry)
XX
DE   HIV-1 subtype C protein fragment #1.
XX
KW   Infection; diagnosis; human; humoral immune response; antiviral;
KW   cellular immune response; vaccine; treatment; gene therapy.
XX
OS   Human immunodeficiency virus 1.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1..3025
FT                   /label= Xaa
FT                   /note= "Xaa represents a stop codon"
XX
PN   DE10056747-A1.
XX
PD   31-MAY-2001.
XX
PF   16-NOV-2000; 2000DE-01056747.
XX
PR   16-NOV-1999;   99DE-01055089.
XX
PA   (SHAO/) SHAO Y.
PA   (GENE-) GENEART GMBH GES ANGEWANDTE BIOTECHNOLOG.
XX
PI   Wagner R,  Wolf H,  Shao Y,  Graf M;
XX
DR   WPI; 2001-336417/36.
DR   N-PSDB; AAH20870.
XX
PT   New nucleic acid sequences from a human immune deficiency virus
PT   intersubtype, useful for treatment, prevention and diagnosis of
PT   infection.
XX
PS   Disclosure; Fig 8A-O; 48pp; German.
XX
CC   This invention describes a novel polynucleotide isolated from human
CC   immunodeficiency virus type 1 subtype C/B' which can be used for the
CC   induction of specific humoral and cellular immune responses. (I) and
CC   polypeptides (II) encoded by them, are useful in pharmaceuticals,
CC   vaccines and diagnostic agents, particularly for treatment or prevention
CC   of human immune deficiency virus-1 (HIV-1) infections, also for rational
CC   design of test or therapeutic reagents, or gene therapy vectors.
CC   Polypeptides, especially antibodies, specifically directed against (II)
CC   are similarly useful as pharmaceutical and diagnostic agents. (I) are
CC   specific for intersubtype C/B' of HIV-1 so are useful in regions
CC   (particularly China and South-East Asia) where this subtype is prevalent.
CC   The products of the invention have antiviral activity. This sequence
CC   represents a protein encoded by the HIV-1 subtype C genome described in
CC   the method of the invention. (Updated on 11-SEP-2003 to standardise OS
CC   field)
XX
SQ   Sequence 3025 AA;

Alignment Scores:
Pred. No.:              1.95e-197      Length:         3025
Score:                  3563.00        Matches:        682
Percent Similarity:     91.65%         Conservative:   42
Best Local Similarity:  86.33%         Mismatches:     57
Query Match:            77.91%         Indels:         9
DB:                     4              Gaps:           5

US-09-610-313B-31 (1-2463) x AAB86169 (1-3025)

Qy         15 GGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAA 74
              ||||||||:::||||||      |||      |||  ||||||    :::|||
Db        416 GlyArgGlyAsnGluProAsnLysGln---CysHisThrAspAlaGluLysGlnPhe*** 434

Qy         75 GGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAA 134
                    :::      :::   ||||||   ||||||      |||      :::
Db        435 ArgLeu***LysAsnCys***MetPheGlnLeuTrpGlnGlyArgAlaHisSerGlnLys 454

Qy        135 CTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAA 194
              |||      |||   ::::::|||||||||||:::    :::          |||:::|||
Db        455 LeuGlnGlyPro***GluLysGlyLeuLeuGluMetTrpLysArgArgThrProAsnGlu 474

Qy        195 GGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAA 254
                 |||:::          |||||| ||||||||||:::||||||   |||||||||::
Db        475 ArgLeuTyr***GluThrGlyGln-PhePheArgGluAsnLeuAlaLeuProGlnGlyAr 494

Qy        255 GGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCA 314
              :||||||||||||   |||||||||   ||||||||||||||||||      ||||||||
Db        494 gAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerProThrArgGlyGluLeuGl 514

Qy        315 GGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCT 368
              ||||      |||   :::|||      |||||||||||||||   ||||||||||||:
Db        514 nValTrpGlyArgAspAsnAsnSerIleSerGluAlaGlyAlaAsnArgGlnGlyThrIl 534

Qy        369 G------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGG 422
              :      ||||||||||||||||||||||||||||||||||||:::||||||:::||
Db        534 eSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGl 554

Qy        423 CGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGAT 482
              |||||||:::||||||||||||||||||||||||||   |||||||||||||||:::::
Db        554 yGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaGlyAspThrValLeuGluAspLe 574

Qy        483 GAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGT 542
              ::::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 uAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysVa 594

Qy        543 GCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCT 602
              |||||||||||:::||||||   ||||||||||||||||||||||||||||||||||||||
Db        594 lArgGlnTyrGluGlnIleProIleGluIleCysGlyHisLysAlaIleGlyThrValLe 614

Qy        603 GATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCAC 662
              |:::||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        614 uValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysTh 634

Qy        663 CCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGA 722
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 rLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAs 654
```

```
Qy        723 CGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCAT 782
Db        654 pGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIl 674

Qy        783 CTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAA 842
Db        674 eCysAspGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAs 694

Qy        843 CACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTT 902
Db        694 nThrProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPh 714

Qy        903 CCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCC 962
Db        714 eArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPr 734

Qy        963 CGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAG 1022
Db        734 oAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSe 754

Qy       1023 CGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAA 1082
Db        754 rIleProLeuTyrGluAspPheArgLysTyrThrAlaPheThrIleProSerArgAsnAs 774

Qy       1083 CGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCC 1142
Db        774 nGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerLe 794

Qy       1143 CAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGA 1202
Db        794 uAlaIlePheGlnSerSerMetThrLysThrLeuGluProPheArgLysGlnAsnProGl 814

Qy       1203 GATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCA 1256
Db        814 yIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGl 834

Qy       1257 GCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCC 1316
Db        834 nHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrPr 854

Qy       1317 CGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGA 1376
Db        854 oAspLysLysHis---LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAs 873

Qy       1377 CAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACAT 1436
Db        873 pLysTrpThrValGlnProThrGlnLeuProGluLysAspSerTrpThrValAsnAspIl 893

Qy       1437 CCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCG 1496
Db        893 eGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValAr 913

Qy       1497 CCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGA 1556
Db        913 gGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGl 933

Qy       1557 GGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGT 1616
Db        933 uGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyVa 953

Qy       1617 GTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTG 1676
Db        953 lTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGluGlnTr 973

Qy       1677 GACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGAT 1736
Db        973 pThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMe 993

Qy       1737 GCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCAT 1796
Db        993 tArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMe 1013

Qy       1797 GGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGAC 1856
Db       1013 tGluGlyIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluTh 1033

Qy       1857 CTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGT 1916
Db       1033 rTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVa 1053

Qy       1917 GAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGC 1976
Db       1053 lAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysAspProIleValGlyVa 1073

Qy       1977 CGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTA 2036
Db       1073 lGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTy 1093

Qy       2037 CGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGAC 2096
Db       1093 rValThrAspArgGlyArgLysLysIleValSerLeuThrGluThrThrAsnGlnLysTh 1113

Qy       2097 CGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGAC 2156
Db       1113 rGluLeuGlnAlaIleCysIleAlaLeuGlnAspSerGlySerGluValAsnIleValTh 1133

Qy       2157 CGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCT 2216
Db       1133 rAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLe 1153

Qy       2217 GGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCC 2276
Db       1153 uValAsnGlnIleIleGluGlnLeuMetLysLysGluArgValTyrLeuSerTrpValPr 1173

Qy       2277 CGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCG 2336
Db       1173 oAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleAr 1193

Qy       2337 CAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db       1193 gLysValLeuPheLeuAspGlyIleAsp 1202
```

```
RESULT 9
AAM48949
ID   AAM48949 standard; protein; 854 AA.
XX
AC   AAM48949;
XX
DT   29-AUG-2003  (revised)
DT   19-APR-2002  (first entry)
XX
DE   HIV-1 subtype C isolate Du151 pol protein.
XX
KW   HIV-1 subtype C; vaccine; HIV infection; AIDS; pol; antiviral.
XX
OS   Human immunodeficiency virus; type I.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 58
FT                   /note= "encoded by TGA"
FT   Misc-difference 72
FT                   /note= "encoded by TAA"
FT   Misc-difference 114
FT                   /note= "encoded by TAA"
FT   Misc-difference 117
FT                   /note= "encoded by TAA"
XX
PN   WO200204494-A2.
XX
PD   17-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-IB001208.
XX
PR   07-JUL-2000; 2000US-0216995P.
PR   10-JUL-2000; 2000ZA-00003437.
```

```
PR   15-SEP-2000; 2000ZA-00004924.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
PA   (UYCA-) UNIV CAPE TOWN.
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Williamson C,  Swanstrom RI,  Morris L,  Karim SA,  Johnston RE;
XX
DR   WPI; 2002-171700/22.
DR   N-PSDB; AAL41592.
XX
PT   Selecting HIV-1 subtype C isolates, which are useful in developing
PT   vaccines against HIV infection, comprises isolating viruses with high
PT   sequence identity to a consensus sequence whose phenotype is associated
PT   with the HIV subtype.
XX
PS   Claim 19; Page 67; 69pp; English.
XX
CC   The present invention relates to a process for the selection of human
CC   immunodeficiency virus (HIV) subtype isolates for use in the development
CC   of a prophylactic and/or therapeutic pharmaceutical composition. The
CC   process involves selecting isolated virus or viruses with a high sequence
CC   identity to a consensus sequence and a phenotype which is associated with
CC   transmission for the particular HIV subtype. The composition can be used
CC   in the production of vaccines against HIV. The present sequence is the
CC   HIV-1 subtype C isolate Du151 pol protein. (Updated on 29-AUG-2003 to
CC   standardise OS field)
XX
SQ   Sequence 854 AA;

Alignment Scores:
Pred. No.:              3.94e-195     Length:        854
Score:                  3521.00       Matches:       666
Percent Similarity:     94.31%        Conservative:  30
Best Local Similarity:  90.24%        Mismatches:    29
Query Match:            77.00%        Indels:        13
DB:                     5             Gaps:          3

US-09-610-313B-31 (1-2463) x AAM48949 (1-854)

Qy        171 CGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTT------ 224
Db        118 ArgGlnGlyPheProSerHisAspValValLysArgArgProValProSerLeuHisAla 137

Qy        225 --------------------CCGCGAGGACCTGGCC---TTCCCCAGGGCAAGGCCCG 260
Db        138 CysArgSerThrLeuGluAspProArgValPro-SerSerPheProGlnGlyProAlaAr 157

Qy        261 CGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCG 320
Db        157 gGlnPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeuGlnValAr 177

Qy        321 CGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCA 380
Db        177 gArgAspAsnProArgSerGluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGl 197

Qy        381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db        197 nIleThrLeuTrpGlnArgProLeuValSerIleLysIleGlyGlyGlnThrArgGluAl 217

Qy        441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db        217 aLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTr 237

Qy        501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db        237 pLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIl 257

Qy        561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db        257 eLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVa 277

Qy        621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db        277 lAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSe 297

Qy        681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db        297 rProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGl 317

Qy        741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db        317 nTrpProLeuThrGluValLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLy 337

Qy        801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db        337 sGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIlePheAlaIl 357

Qy        861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db        357 eLysLysGluAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysAr 377

Qy        921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db        377 gThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLy 397

Qy        981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db        397 sLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGl 417

Qy       1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db        417 yPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleAr 437

Qy       1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db        437 gTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSe 457

Qy       1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG-- 1218
Db        457 rMetThrLysIleLeuGluProPheArgAlaLysAsnProGluIleValIleTyrGlnTy 477

Qy       1219 ----GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db        477 rMetAlaAlaLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGl 497

Qy       1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db        497 uGluLeuArgGluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLy 517

Qy       1335 GGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1394
Db        517 sGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnPr 537

Qy       1395 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db        537 oIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLy 557

Qy       1455 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1514
Db        557 sLeuAsnTrpThrSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLe 577

Qy       1515 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1574
Db        577 uArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLe 597

Qy       1575 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1634
Db        597 uAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLy 617

Qy       1635 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1694
Db        617 sAspLeuIleAlaGluIleGlnLysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGl 637

Qy       1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1754
```

```
Db        637 nGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysArgArgThrThrHisThrAs 657

Qy       1755 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814
Db        657 nAspValLysGlnLeuThrGluAlaValGlnLysIleSerLeuGluSerIleValThrTr 677

Qy       1815 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db        677 pGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpTh 697

Qy       1875 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1934
Db        697 rAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVa 717

Qy       1935 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994
Db        717 lLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPheTyrValAs 737

Qy       1995 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2054
Db        737 pGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyAr 757

Qy       2055 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2114
Db        757 gGlnLysIleValThrLeuSerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGl 777

Qy       2115 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174
Db        777 nLeuAlaLeuGlnAspSerGluSerGluValAsnIleValThrAspSerGlnTyrAlaLe 797

Qy       2175 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2234
Db        797 uGlyIleIleGlnAlaGlnProAspArgSerGluSerGluLeuValAsnGlnIleIleGl 817

Qy       2235 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2294
Db        817 uGlnLeuIleLysLysGluArgAlaTyrLeuSerTrpValProAlaHisLysGlyIleGl 837

Qy       2295 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
Db        837 yGlyAspGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 854
```

RESULT 10
AAB69282
ID   AAB69282 standard; protein; 1000 AA.
XX
AC   AAB69282;
XX
DT   12-SEP-2003 (revised)
DT   20-APR-2001 (first entry)
XX
DE   HIV-1 non-subtype B clone 92RW009-6 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999; 99WO-US024837.
XX
PR   02-NOV-1998; 98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH, Shaw GM, Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1000 AA;

Alignment Scores:

Pred. No.:	6.5e-195	Length:	1000
Score:	3517.50	Matches:	660
Percent Similarity:	95.96%	Conservative:	29
Best Local Similarity:	91.92%	Mismatches:	26
Query Match:	76.92%	Indels:	3
DB:	3	Gaps:	2

US-09-610-313B-31 (1-2463) x AAB69282 (1-1000)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArgLysPheSerProGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
Db         21 ThrGlyAlaAsnSerProThrSerArgGluLeuTrpAsnGlyGlyArgAspSerLeuSer 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAG 396
Db         41 SerGluThrGlyAlaGluArgGlnGlyThrPheAsnPheProGlnIleThrLeuTrpGln 60

Qy        397 CGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGC 456
Db         61 ArgProLeuValThrValLysIleGlyGlyGlnLeuArgGluAlaLeuLeuAspThrGly 80

Qy        457 GCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATC 516
Db         81 AlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIle 100

Qy        517 GGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGC 576
Db        101 GlyGlyIleGlyGlyPheIleLysValLysGlnTyrAspGlnIleLeuIleGluIleCys 120

Qy        577 GGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGC 636
Db        121 GlyLysLysAlaIleGlyThrValLeuValGlyProThrSerValAsnIleIleGlyArg 140

Qy        637 AACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTG 696
Db        141 AsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal 160

Qy        697 CCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAG 756
Db        161 ProValAlaLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 180

Qy        757 GAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACC 816
Db        181 GluLysIleLysAlaLeuArgGluIleCysThrGluMetGluLysGluGlyLysIleSer 200

Qy        817 AAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGC 876
Db        201 LysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 220

Qy        877 ACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936
Db        221 ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp 240
```

```
Qy        937 GAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTG 996
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVal 260

Qy        997 CTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACC 1056
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        261 LeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThr 280

Qy       1057 GCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTG 1116
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 300

Qy       1117 CTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTG 1176
              |||||||||||||||||||||||||||:::||||||||:::|||||||||||||||||||
Db        301 LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAsnSerMetThrLysIleLeu 320

Qy       1177 GAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTAC 1230
              ||||||||||||||:::|||   ||||||||||||||||           ||||||
Db        321 GluProPheArgAlaGlnAsnGlnGluIleValIleTyrGlnTyrMetAspAspLeuTyr 340

Qy       1231 GTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCAC 1290
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        341 ValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHis 360

Qy       1291 CTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG 1350
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 380

Qy       1351 TGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAG 1410
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        381 TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGlu 400

Qy       1411 AAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGC 1470
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 420

Qy       1471 CAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCC 1530
              ||||||||||||:::||||||||||||||||||||||||||||||||      ||||||
Db        421 GlnIleTyrProGlyValLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAla 440

Qy       1531 CTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAG 1590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 LeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 460

Qy       1591 ATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAG 1650
              ||||||:::|||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        461 IleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGlu 480

Qy       1651 ATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAAC 1710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 IleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 500

Qy       1711 CTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTG 1770
              |||||||||||||||||||||||||     ||||||||||||||||||||||||||||||
Db        501 LeuLysThrGlyLysTyrAlaLysArgArgThrAlaHisThrAsnAspValLysGlnLeu 520

Qy       1771 ACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAG 1830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 ThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrpGlyLysThrProLys 540

Qy       1831 TTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCC 1890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAla 560

Qy       1891 ACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAG 1950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 580

Qy       1951 CTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGC 2010
              ||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        581 LeuGluLysGluProIleLeuGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg 600

Qy       2011 GAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGC 2070
              |||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db        601 GluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSer 620

Qy       2071 CTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGAC 2130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 LeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAsp 640

Qy       2131 AGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCC 2190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 SerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 660

Qy       2191 CAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAG 2250
              ||||||||    |||||||||||    |||||||||||||||||||||||||||||||||
Db        661 GlnProAspSerSerGluSerGluAlaValAsnGlnIleIleGluGlnLeuIleLysLys 680

Qy       2251 GAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATC 2310
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        681 GluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 700

Qy       2311 GACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              |||||||||||||||      |||||||:::||||||||||||||||||||||
Db        701 AspLysLeuValSerSerGlyIleArgArgValLeuPheLeuAspGlyIleAsp 718
```

RESULT 11
AAW72993

```
ID   AAW72993 standard; protein; 1002 AA.
XX
AC   AAW72993;
XX
DT   25-MAR-2003  (revised)
DT   15-FEB-1999  (first entry)
XX
DE   HIV isolate LAV.MAL pol protein.
XX
KW   LAV.MAL; HIV; human immunodeficiency virus; diagnosis; vaccine; AIDS;
KW   pol.
XX
OS   Human T-lymphotropic virus.
XX
PN   US5824482-A.
XX
PD   20-OCT-1998.
XX
PF   06-JUN-1995;   95US-00471474.
XX
PR   23-JUN-1986;   86FR-00040138.
PR   13-APR-1987;   87US-00038330.
PR   19-FEB-1991;   91US-00656797.
PR   10-DEC-1992;   92US-00988530.
PR   18-NOV-1993;   93US-00154397.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Wain-Hobson S,  Sonigo P,  Alizon M,  Montagnier L;
XX
DR   WPI; 1998-582548/49.
DR   N-PSDB; AAV63467.
XX
PT   Human immunodeficiency virus isolate LAV(MAL) - and method for detecting
PT   anti-HIV antibodies.
XX
PS   Disclosure; Fig 7B-E; 47pp; English.
XX
CC   This is the amino acid sequence of the pol protein of lymphadenopathy
CC   associated virus LAV.MAL (CNCM I-641), a new virus isolate from Zaire
CC   that is responsible for diseases clinically related to AIDS. The sequence
CC   was deduced from an open reading frame (ORF) of the LAV.MAL genome (see
CC   AAV63467). 7 ORFs (see AAW72992-98) were identified. Specific peptides of
CC   the envelope glycoprotein can be used as antigens in a claimed method for
CC   the in vitro detection of an antibody directed against LAV. The method is
CC   useful for the diagnosis of AIDS or pre-AIDS, or to detect antibodies in
CC   patients, asymptomatic carriers and in blood-related products. LAV viral
```

```
CC   antigens are also useful in vaccines. (Updated on 25-MAR-2003 to correct
CC   PR field.)
XX
SQ   Sequence 1002 AA;

Alignment Scores:
Pred. No.:              7.14e-194     Length:         1002
Score:                  3499.50       Matches:        653
Percent Similarity:     96.11%        Conservative:   39
Best Local Similarity:  90.69%        Mismatches:     23
Query Match:            76.53%        Indels:         5
DB:                     2             Gaps:           3

US-09-610-313B-31 (1-2463) x AAW72993 (1-1002)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
              ||||||||||||||||||||||||||||||||:::|||         ||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACCCTG 390
              ||||||   |||||||||||||||||          :::    :::||||||||||||||
Db         41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPheSerPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              |||||||||||:::|||:::::::::||||||||||||:::|||||||||||||||||||
Db         61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              |||||||||||||||||||||||||||||||::::::|||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              |||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              ||||||||||||||||||||||||||    ||||||::::::||||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              |||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
              ||||||||||||||||    :::||||||||||||||||||||||||
Db        321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy       1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy       1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
              :::||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy       1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400

Qy       1405 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1464
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy       1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1524
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440

Qy       1525 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1584
              ||||||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db        441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460

Qy       1585 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1644
              |||||||||||:::||||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1645 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1704
              ||||||||||||||||||||       |||||||||||||||||||||||||    :::
Db        481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500

Qy       1705 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1764
              |||||||||||||||||||||||||||||:::::::::::||||||||||||||||||||
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520

Qy       1765 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1824
              |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540

Qy       1825 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884
              ||||||||||||||||||||||||||||||||||||||    ||||||||:::|||||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy       1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1945 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2004
              ||||||||||||   |||||||||:::|||||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       2005 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2064
              ||||||||||||||    ||||||||||||||||||||||||||||||||||||||:::
Db        601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2065 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2124
              |||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy       2125 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660
```

```
Qy       2185 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2244
               ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

Qy       2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2304
              :::|||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2305 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              |||:::||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 12
AAO30963
ID AAO30963 standard; protein; 1003 AA.
XX
AC AAO30963;
XX
DT 06-OCT-2003 (first entry)
XX
DE HIV pol protein.
XX
KW Human immunodeficiency virus; HIV; pol; IV9; immunotherapy; infection;
KW epitope; vaccine; enzyme.
XX
OS Human immunodeficiency virus.
XX
PN WO2003051285-A2.
XX
PD 26-JUN-2003.
XX
PF 29-OCT-2002; 2002WO-US034620.
XX
PR 29-OCT-2001; 2001US-0345116P.
XX
PA (GENZ) GENZYME CORP.
PA (GEHO) GEN HOSPITAL CORP DBA MASSACHUSETTS GEN.
XX
PI Nicolette CA, Walker BD;
XX
DR WPI; 2003-541603/51.
XX
PT Novel synthetic therapeutic peptide compounds useful for inducing an
PT immune response in a mammal, specifically cross-react with HIV IV9
PT epitopes.
XX
PS Claim 8; Page 65-67; 67pp; English.
XX
CC The invention relates to synthetic therapeutic peptide compounds that
CC specifically cross-react with HIV (human immunodeficiency virus) pol
CC protein IV9 epitopes. The invention is useful for treating HIV infection.
CC The invention is also used in immunotherapy and as vaccines. The present
CC sequence is HIV pol protein
XX
SQ Sequence 1003 AA;

```
Alignment Scores:
Pred. No.:              9.96e-194      Length:         1003
Score:                  3497.00        Matches:        655
Percent Similarity:     95.56%         Conservative:   34
Best Local Similarity:  90.85%         Mismatches:     26
Query Match:            76.47%         Indels:         6
DB:                     6              Gaps:           3
```

US-09-610-313B-31 (1-2463) x AAO30963 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||   |||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                 ||||||||||||||||||   ||||||||||||||      |||   :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||:::||||||||||||:::      ||||||||||:::|||
Db         41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              ||||||||||||||||||:::|||||:::|||||||||:::|||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              |||||||||||   |||||||||||||||||||:::|||||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              |||||||||||:::|||||||||:::||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||||         ||||||  ||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              ||||||||||||||||||||   :::||||||:::|||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 |||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
```

```
Qy       1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                 ||||||||||||:::::::::||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              :::|||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              ||||||||||||||||||||||||||||||:::||||||   ||||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              |||||||||||||||||||||||||||      |||||||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              ||||||||||||||||||:::|||||||||||||||||||||:::|||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::|||:::||||||:::|||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::|||||||||||||||   |||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              |||
Db        721 Asp 721
```

RESULT 13
ADN36413
ID   ADN36413 standard; protein; 1003 AA.
XX
AC   ADN36413;
XX
DT   15-JUL-2004 (first entry)
XX
DE   HIV protein for anti-HIV vaccine.
XX
KW   anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO2004035006-A2.
XX
PD   29-APR-2004.
XX
PF   17-OCT-2003; 2003WO-US033112.
XX
PR   18-OCT-2002; 2002US-0419465P.
XX
PA   (AARO-) AARON DIAMOND AIDS RES CENT.
XX
PI   Huang Y, Ho DD, Chen Z;
XX
DR   WPI; 2004-348328/32.
DR   N-PSDB; ADN36412.
XX
PT   Nucleic acid vector comprising at least one HIV sequence operably linked
PT   to a promoter and encoding a protein that does not assemble into viral
PT   particles, useful in immunizing a subject against HIV infection.
XX
PS   Disclosure; SEQ ID NO 27; 166pp; English.
XX
CC   The invention relates to a nucleic acid vector comprising at least one
CC   HIV sequence operably linked to a promoter and encoding a protein that
CC   does not assemble into viral particles. The nucleic acid vector is useful
CC   in immunizing a subject against HIV infection. This sequence corresponds
CC   to a nucleic acid used in the invention.
XX
SQ   Sequence 1003 AA;

```
Alignment Scores:
Pred. No.:              9.96e-194     Length:        1003
Score:                  3497.00       Matches:       655
Percent Similarity:     95.56%        Conservative:  34
Best Local Similarity:  90.85%        Mismatches:    26
Query Match:            76.47%        Indels:        6
DB:                     8             Gaps:          3

US-09-610-313B-31 (1-2463) x ADN36413 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||   ||||||||||||||||||   ||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                 ||||||||||||||||||   |||||||||||||||      |||   :::|||   
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||||||:::||||||||||||:::      ||||||||||||:::|||
Db         41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||||:::||||||:::|||||||||:::|||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
```

```
Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db    321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy   1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db    341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy   1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db    361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy   1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
Db    381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy   1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
Db    401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy   1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
Db    421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy   1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
Db    441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy   1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
Db    461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy   1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
Db    481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy   1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
Db    501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy   1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
Db    521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy   1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
Db    541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy   1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
Db    561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy   1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
Db    581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy   2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
Db    601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy   2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
Db    621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy   2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
Db    641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy   2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
Db    661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy   2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
Db    681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy   2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
Db    701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy   2362 GAT 2364
Db    721 Asp 721
```

RESULT 14
AAP81861

```
ID   AAP81861 standard; protein; 1002 AA.
XX
AC   AAP81861;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   16-DEC-1990  (first entry)
XX
DE   Sequence encoded by LAV MA L POL gene.
XX
KW   HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.
XX
OS   Lymphadenopathy-associated virus.
XX
PN   WO8707906-A.
XX
PD   30-DEC-1987.
XX
PF   22-JUN-1987;   87WO-EP000326.
XX
```

```
PR   23-JUN-1986;   86EP-00401380.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Alizon M,  Sonigo P,  Wainhobson S,  Montagnier L,  Alizon M;
PI   Sonigo P,  Wainhobson S,  Montagnier L;
XX
DR   WPI; 1988-014396/02.
DR   N-PSDB; AAN80437.
XX
PT   New variants of lymphadenopathy associated virus (LAV) - used for prodn.
PT   of DNA, antigens and antibodies used in diagnosis of AIDS and pre-AIDS.
XX
PS   Claim 8; Fig 8A-8I; 72pp; English.
XX
CC   LAV EL I (AAN80436) and LAV MA L (AAN80437) were isolated from the
CC   peripheral blood lymphocytes of patients. Different AIDS virus isolates
CC   concerned are designated by 3 letters of the patients name. Stable probes
CC   including the DNA sequences can be used for detection of the new LAV
CC   viruses or related viruses or DNA proviruses in eg. biological samples.
CC   The proteins or peptides can be used for detection of antibodies induced
CC   in vivo and present in biological fluids. The DNA can also be used for
CC   the expression of LAV viral antigens for the prodn. of a vaccine against
CC   LAV. The polypeptides can also be used for the prodn. of antibodies for
CC   the detection of proteins related to the LAV viruses, partic. for
CC   diagnosis of AIDS or pre-AIDS. (Updated on 25-MAR-2003 to correct PI
CC   field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 1002 AA;

Alignment Scores:
Pred. No.:              1.39e-193     Length:         1002
Score:                  3494.50       Matches:        652
Percent Similarity:     95.97%        Conservative:   39
Best Local Similarity:  90.56%        Mismatches:     24
Query Match:            76.42%        Indels:         5
DB:                     1             Gaps:           3

US-09-610-313B-31 (1-2463) x AAP81861 (1-1002)

Qy      220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db        1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy      280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
Db       21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40

Qy      337 AGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACCCTG 390
Db       41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPheSerPheProGlnIleThrLeu 60

Qy      391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
Db       61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy      451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
Db       81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy      511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
Db      101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy      571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
Db      121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140

Qy      631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
Db      141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy      691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
Db      161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180

Qy      751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
Db      181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200

Qy      811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
Db      201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy      871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
Db      221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy      931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
Db      241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy      991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
Db      261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy     1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
Db      281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy     1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
Db      301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy     1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
Db      321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy     1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
Db      341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy     1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
Db      361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy     1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404
Db      381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400

Qy     1405 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1464
Db      401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy     1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1524
Db      421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440

Qy     1525 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1584
Db      441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuProGluAsn 460

Qy     1585 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1644
Db      461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy     1645 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1704
Db      481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500

Qy     1705 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1764
Db      501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520

Qy     1765 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1824
```

```
Db        521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540

Qy       1825 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884
              |||||||||||||||||||||||||||||||||||||   ||||||||:::||||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy       1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1945 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2004
              ||||||||||||   |||||||||:::||||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       2005 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2064
              ||||||||||||||   |||||||||||||||||||||||||||||||||||||||:::
Db        601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2065 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2124
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy       2125 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2185 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2244
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

Qy       2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2304
              :::|||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2305 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              |||:::||||||||||||||   |||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 15
ABR55489
```
ID   ABR55489 standard; protein; 1003 AA.
XX
AC   ABR55489;
XX
DT   11-AUG-2003  (first entry)
XX
DE   Amino acid sequence of a HIV pol protein.
XX
KW   Polymorphic marker; host response; HIV; AIDS related virus; vaccine;
KW   hepatitis related virus; HCV; HBV; drug resistance; pol.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003035097-A1.
XX
PD   01-MAY-2003.
XX
PF   23-OCT-2002; 2002WO-AU001450.
XX
PR   23-OCT-2001; 2001AU-00008425.
XX
PA   (EPIP-) EPIPOP PTY LTD.
XX
PI   Mallal S;
XX
DR   WPI; 2003-449231/42.
XX
PT   Determining the influence of variation in host genes on the selection of
PT   microorganisms with protein substitutions, comprises typing individuals
PT   of a cohort infected with a microorganism for an intrinsic polymorphic
PT   marker.
XX
PS   Claim 22; Page 88-89; 157pp; English.
XX
CC   The specification describes a method of determining the influence of
CC   variation in host genes on selection of microorganisms with protein
CC   substitutions. The method comprises typing all individuals of a
CC   population of patients infected with a microorganism for at least one
CC   selected intrinsic polymorphic marker involved in the host response to
CC   the presence of the microorganism. The method is useful for examining
CC   selective pressures confronting a wide range of organisms that exhibit
CC   pathogenic traits in a host, such as bacteria, fungi, mycobacterium,
CC   viruses and virus-like particles; for examining microorganisms that have
CC   adapted to evolve rapidly, including HIV and AIDS related viruses and the
CC   hepatitis related viruses such as HCV and HBV. The method is useful for
CC   designing a vaccine to prevent or delay the emergence of drug resistance
CC   in patients treated with a particular drug specific for a microorganism
CC   where the drug affects the replication of the microorganism at the
CC   nucleotide or amino acid level. The present sequence represents a HIV pol
CC   protein, which is expected to provide optimal cytotoxic T lymphocyte
CC   (CTL) induced therapeutic protection to the cohort examined in that study
XX
SQ   Sequence 1003 AA;

Alignment Scores:
Pred. No.:              1.48e-193      Length:        1003
Score:                  3494.00        Matches:       654
Percent Similarity:     95.70%         Conservative:  36
Best Local Similarity:  90.71%         Mismatches:    25
Query Match:            76.40%         Indels:        6
DB:                     6              Gaps:          3

US-09-610-313B-31 (1-2463) x ABR55489 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||:::|||||||||||||||||||||||||||||||||   ||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGAC------AACCCC 333
                 ||||||||||||||||   |||||||||||||   |||:::      |||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGluAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 |||||||||||||:::||||||||||||:::      :::||||||||||||
Db         41 ThrSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              ||||||||||||||||||||:::||||||:::||||||||:::|||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||:::|||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleIleIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              ||||||||||||   ||||||||||||||||||||:::||||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              |||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              |||||||||||||||||||||||||||       ||||||   ||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              ||||||||||||||||||||     :::||||||:::||||||||||||||      
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 |||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::|||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
              ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                 ||||||||||||:::::::::||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              :::|||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              ||||||||||||||||||||||||||||||:::||||||   ||||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
              ||||||||||||:::||||||||||||||||||||||||||   |||||||||:::||||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::||||||||||||:::|||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::|||||||||||||||   |||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              |||
Db        721 Asp 721
```

Search completed: June 2, 2005, 04:42:47
Job time : 231.833 secs

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:          June  2, 2005, 02:53:39 ; Search time 40.8333 Seconds
                                           (without alignments)
                                           11607.276 Million cell updates/sec

Title:           US-09-610-313B-31
Perfect score:   4573
Sequence:        1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2463

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:      566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155137_15159/app_query.fasta_1.7893
-DB=PIR_79 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_189_@runat_31052005_155137_15159 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       PIR_79:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3505.5	76.7	1002	2	S54378	pol polyprotein -
2	3497.5	76.5	1002	1	GNLJND	HIV-1 retropepsin
3	3492	76.4	1003	1	GNVWLV	HIV-1 retropepsin
4	3475	76.0	1015	1	GNVWH3	HIV-1 retropepsin
5	3453	75.5	1003	2	T09440	pol polyprotein -
6	3450	75.4	1003	1	B44001	HIV-1 retropepsin
7	3449	75.4	1003	1	GNVWA2	HIV-1 retropepsin
8	3439	75.2	1012	1	GNVWVL	HIV-1 retropepsin
9	3211	70.2	1027	1	GNLJSI	HIV-1 retropepsin
10	3100	67.8	912	2	S33980	pol polyprotein -
11	3094	67.7	902	2	T01668	pol polyprotein -
12	2771	60.6	559	2	B47175	reverse transcript
13	2751	60.2	559	2	A47175	reverse transcript
14	2350	51.4	1039	2	S46347	pol polyprotein -
15	2209	48.3	1055	2	S53092	pol polyprotein -
16	2205.5	48.2	1055	1	GNLJST	HIV-1 retropepsin
17	2196.5	48.0	1056	1	GNLJG3	HIV-1 retropepsin
18	2196	48.0	1036	1	GNLJG2	HIV-1 retropepsin
19	2188	47.8	1034	1	GNLJCA	HIV-1 retropepsin
20	2187.5	47.8	1035	1	GNLJGG	HIV-1 retropepsin
21	2182.5	47.7	1019	2	T11560	pol polyprotein -
22	2168.5	47.4	1009	2	S28081	pol polyprotein -
23	2150.5	47.0	1032	2	S12153	pol polyprotein -
24	2073	45.3	1058	2	S08436	pol polyprotein -
25	2053	44.9	1061	1	GNLJG4	HIV-1 retropepsin
26	2019.5	44.2	1054	1	GNLJG5	HIV-1 retropepsin
27	1963.5	42.9	656	2	S30484	pol polyprotein -
28	1934.5	42.3	656	2	S30483	pol polyprotein -
29	1390	30.4	1124	2	S23820	pol polyprotein -
30	1387	30.3	1124	1	GNLJFP	HIV-1 retropepsin
31	1344	29.4	1124	2	B45557	HIV-1 retropepsin
32	1148.5	25.1	1109	1	B45345	HIV-1 retropepsin
33	1134	24.8	1146	1	GNLJ22	HIV-1 retropepsin
34	1134	24.8	1146	1	GNLJEW	HIV-1 retropepsin
35	1131.5	24.7	1087	2	JQ1162	Pol protein - Maed
36	1128.5	24.7	1086	1	B46335	HIV-1 retropepsin
37	1121	24.5	1101	1	B45390	HIV-1 retropepsin
38	1117	24.4	1101	1	GNLJVS	HIV-1 retropepsin
39	1114.5	24.4	1145	1	GNLJEV	HIV-1 retropepsin
40	1090	23.8	216	2	S43127	reverse transcript
41	1071	23.4	219	2	S32152	RNA-directed DNA p
42	1067	23.3	219	2	S32118	RNA-directed DNA p
43	1067	23.3	219	2	S32139	RNA-directed DNA p
44	1065	23.3	219	2	S32089	RNA-directed DNA p
45	1065	23.3	219	2	S32160	RNA-directed DNA p

ALIGNMENTS

```
RESULT 1
S54378
pol polyprotein - human immunodeficiency virus type 1 (fragment)
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C;Accession: S54378
R;Theodore, T.; Buckler-White, A.J.
submitted to the EMBL Data Library, July 1989
A;Reference number: S54377
A;Accession: S54378
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1-1002 <THE>
A;Cross-references: UNIPROT:P12499; EMBL:M22639; NID:g329377; PIDN:AAA45366.1; PID:g3293
C;Superfamily: pol polyprotein
C;Keywords: polyprotein

Alignment Scores:
Pred. No.:              2.59e-155     Length:         1002
Score:                  3505.50       Matches:        655
Percent Similarity:     95.69%        Conservative:   34
Best Local Similarity:  90.97%        Mismatches:     26
Query Match:            76.66%        Indels:         5
DB:                     2             Gaps:           3

US-09-610-313B-31 (1-2463) x S54378 (1-1002)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||||||||||||||||||||  |||    |||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
                 ||||||||||||||||||||||||||||:::|||   |||    |||||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTC------CCCCAGATCACCCTG 390
              |||||   |||||||||||||||||||::::::|||         ||||||||||||||
Db         41 SerGluThrGlyAlaGluArgGlnGlyThrValSerPheAsnCysProGlnIleThrLeu 60
```

```
Qy       391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
             |||||||||||||||:::||||||:::|||||||||:::|||||||||||||||||||||
Db        61 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy       451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
             ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100

Qy       511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy       571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
             |||||||||   ||||||||||||||||||:::|||||||||||||||||||||||||||
Db       121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy       631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
             |||||||||:::|||||||||:::||||||||||||||||||||||||||||||||||||
Db       141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy       691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy       751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
             |||||||||||||||||||||||||||   ||||||   |||||||||||||||||||||
Db       181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200

Qy       811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
             |||:::::::::||||||||||||||||||||||||:::|||||||||||||||||||||
Db       201 IleSerArgValGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220

Qy       871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy       931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy       991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
             ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLys 280

Qy      1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy      1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
             |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db       301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy      1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
             ||||||||||||||||||   :::||||||||||||||||||||||||            
Db       321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy      1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
             |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db       341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy      1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
             :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy      1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404
             ||||||||||||||||||||||||||||||||||||||||||||||||   |||:::|||
Db       381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400

Qy      1405 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy      1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1524
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||   
Db       421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr 440

Qy      1525 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1584
             |||||||||:::::::::||||||||||||||||||||||||||||||||||||||||||
Db       441 LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460

Qy      1585 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1644
             |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||:::
Db       461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy      1645 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1704
             ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db       481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy      1705 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1764
             ||||||||||||||||||||||||:::|||||||||   |||||||||||||||||||||
Db       501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520

Qy      1765 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1824
             ||||||   |||   |||||||||:::                                 
Db       521 GlnLeuAlaGluValValGlnLysIleSerThrGluSerIleValIleTrpGlyLysThr 540

Qy      1825 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884
             ||||||||||||||||||||||||||||||||||||||||||||||||   :::||||||
Db       541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpValGluTyrTrp 560

Qy      1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy      1945 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy      2005 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2064
             |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||:::
Db       601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy      2065 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2124
             |||   |||:::||||||||||||||||||||||||||||||||||||   |||||||||
Db       621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640

Qy      2125 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2184
             ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db       641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy      2185 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2244
             |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db       661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680

Qy      2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2304
             ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db       681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy      2305 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
             |||:::|||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db       701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 2
GNLJND
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate NDK)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA polym
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: JQ0067
R;Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.;
Gene 81, 275-284, 1989
A;Title: Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunoc
A;Reference number: JQ0065; MUID:90034200; PMID:2806917
A;Accession: JQ0067

```
A;Molecule type: DNA
A;Residues: 1-1002 <SPI>
A;Cross-references: UNIPROT:P18802; GB:M27323; NID:g328154; PIDN:AAA44869.1; PID:g328158
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;56-154/Product: retropepsin #status predicted <RTP>
F;80/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:
Pred. No.:              6.08e-155     Length:        1002
Score:                  3497.50       Matches:       655
Percent Similarity:     95.69%        Conservative:  34
Best Local Similarity:  90.97%        Mismatches:    26
Query Match:            76.48%        Indels:        5
DB:                     1             Gaps:          3

US-09-610-313B-31 (1-2463) x GNLJND (1-1002)

Qy     220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
           ||||||||||||||||||||||||||||||||||||||    ||||||    ||||||||
Db       1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaGlyGluPheSerSerGluGln 20

Qy     280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
              |||||||||||||||||||||||||||:::|||      |||||||||||
Db      21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspAsnProLeu 40

Qy     337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTG 390
           ||||||  ||||||||||||||||||||:::        :::|||||||||||||||||
Db      41 SerGluThrGlyAlaGluArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60

Qy     391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
           ||||||||||||||||:::|||||:::|||||||||:::||||||||||||||||||
Db      61 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy     451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
           ||||||||||||||||||||||||||||||:::::::||||||||||||||||||||||
Db      81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy     511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy     571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
           |||||||||   ||||||:::|||||||||||:::||||||||||||||||||||||||
Db     121 IleCysGlyTyrLysAlaMetGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy     631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
           ||||||||:::||||||||:::||||||||||||||||||||||||||||||||||||
Db     141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy     691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy     751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
           ||||||||||||||||||||||||||    ||||||    ||||||||||||||||||
Db     181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200

Qy     811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
           |||::::::|||||||||||||||||||||||||||||:::||||||||||||||||||
Db     201 IleSerArgIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220

Qy     871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy     931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy     991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy    1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy    1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
           |||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db     301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy    1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
           |||||||||||||||||     :::||||||||||||||||||||||||
Db     321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy    1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
           ||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db     341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy    1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
           :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy    1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404
           ||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db     381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleAsnLeu 400

Qy    1405 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1464
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy    1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1524
           ||||||||||||||   ||||||||||||:::||||||||||||||||||||||||
Db     421 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 440

Qy    1525 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1584
           |||||||||||::::::|||||||||||||||||||||||||||||||||||||||||||
Db     441 LysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460

Qy    1585 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1644
           |||||||||||:::|||||||||||||||||||||||||||||||||||||||||:::
Db     461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy    1645 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1704
           ||||||:::|||||||||||||        ||||||||||||||||||||||||||||||
Db     481 AlaGluLeuGlnLysGlnGlyAspGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy    1705 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1764
           ||||||||||||||||||||||||:::    |||   ||||||||||||||||||||||
Db     501 LysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspValLys 520

Qy    1765 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1824
           ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db     521 GlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 540

Qy    1825 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884
           |||||||||:::|||||||||||||||||||||||||||||||||||    :::||||||
Db     541 ProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpIleGluTyrTrp 560

Qy    1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1944
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy    1945 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy    2005 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2064
           ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||:::
Db     601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy    2065 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2124
```

```
                |||      |||:::|||||||||||||||||||||||||||||||    |||||||||
Db        621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640

Qy       2125 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2184
              |||||||||||    ||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2185 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2244
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680

Qy       2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2304
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2305 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              |||:::|||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 3
GNVWLV
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LAV-1a)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 03-Jun-2002
C;Accession: A03966
R;Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.
Cell 40, 9-17, 1985
A;Title: Nucleotide sequence of the AIDS virus, LAV.
A;Reference number: A90866; MUID:85099333; PMID:2981635
A;Accession: A03966
A;Molecule type: DNA
A;Residues: 1-1003 <WAI>
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;57-155/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status experimental

```
Alignment Scores:
Pred. No.:              1.09e-154     Length:          1003
Score:                  3492.00       Matches:         655
Percent Similarity:     95.42%        Conservative:    33
Best Local Similarity:  90.85%        Mismatches:      27
Query Match:            76.36%        Indels:          6
DB:                     1             Gaps:            3

US-09-610-313B-31 (1-2463) x GNVWLV (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                 ||||||||||||||||||   ||||||||||||||         |||    :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 |||||||||||||:::|||||||||||:::         ||||||||||||||||
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||||:::||||||:::||||||||:::|||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              |||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              |||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              |||||||||||    |||||||||||||||||||||:::|||||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||         ||||||    |||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              ||||||||||||||||||||||      :::||||||:::|||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 |||||||||||||||||||||||||||||||||||||      ||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                 ||||||||||||:::::::|||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
```

```
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              :::||||||||||||||||||         ||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              |||||||||||||||||||||||||||||:::   |||   ||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              |||||||||||||||||||||||||||||       ||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
              |||||||||||:::||||||||||||||||||||||||||||||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              |||:::||||||||||:::|||||||||||||||||||||:::|||||||||||||||||
Db        601 AlaSerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::|||:::|||||:::||||||||||||||||||||||||||||||   ||||||
Db        621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::||||||||||||||||   ||||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              |||
Db        721 Asp 721
```

RESULT 4

GNVWH3

HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate HTLV-III
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C;Accession: A03965
R;Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K.J.; Starcich, B.; Josephs, S.F.; Dora
nberger, J.A.; Papas, T.S.; Ghrayeb, J.; Chang, N.T.; Gallo, R.C.; Wong-Staal, F.
Nature 313, 277-284, 1985
A;Title: Complete nucleotide sequence of the AIDS virus, HTLV-III.
A;Reference number: A93353; MUID:85111123; PMID:2578615
A;Accession: A03965
A;Molecule type: DNA
A;Residues: 1-1015 <RAT>
A;Cross-references: UNIPROT:P03366; GB:M15654; GB:K02008; GB:K02009; GB:K02010; NID:g326
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;69-167/Product: retropepsin #status predicted <RTP>
F;93/Active site: Asp (shared with dimeric partner) #status experimental

```
Alignment Scores:
Pred. No.:                6.73e-154      Length:         1015
Score:                    3475.00        Matches:        654
Percent Similarity:       93.72%         Conservative:   33
Best Local Similarity:    89.22%         Mismatches:     28
Query Match:              75.99%         Indels:         18
DB:                       1              Gaps:           4

US-09-610-313B-31 (1-2463) x GNVWH3 (1-1015)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
              ||||||||||||||||||||||||   |||||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
                                      |||||||||   ||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||   :::|||      ||||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      ||||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::|||||||||||||||||||||||||||||||||:::|||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
              |||      ||||||   |||||||||||||||||||||||:::||||||||||||||||
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
```

```
Db     281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy    1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db     301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy    1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db     321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy    1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db     341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy    1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db     361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy    1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
Db     381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy    1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
Db     401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy    1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
Db     421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy    1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
Db     441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy    1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
Db     461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy    1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
Db     481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy    1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
Db     501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy    1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
Db     521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy    1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
Db     541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy    1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
Db     561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy    1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
Db     581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy    1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
Db     601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy    2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
Db     621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy    2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
Db     641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy    2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
Db     661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy    2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
Db     681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy    2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
Db     701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy    2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db     721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 5
T09440
pol polyprotein - human immunodeficiency virus type 1 (strain JRFL) (fragment)
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T09440
R;Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie,
submitted to the EMBL Data Library, July 1996
A;Reference number: Z16673
A;Accession: T09440
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1003 <PAN>
A;Cross-references: UNIPROT:Q75755; EMBL:U63632; NID:g1465777; PID:g1465779
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	7.03e-153	Length:	1003
Score:	3453.00	Matches:	646
Percent Similarity:	95.01%	Conservative:	39
Best Local Similarity:	89.60%	Mismatches:	30
Query Match:	75.51%	Indels:	6
DB:	2	Gaps:	3

US-09-610-313B-31 (1-2463) x T09440 (1-1003)

```
Qy     220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db       1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy     280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db      21 ThrArgAlaAsnSerProThrArgLysGluLeuGlnValTrpGlyArgAspSerAsnSer 40

Qy     334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACC 387
Db      41 LeuSerGluAlaGlyAlaAspArgGlnGlyAsnValSerPheAsnPheProGlnIleThr 60

Qy     388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db      61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy     448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db      81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy     508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db     101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy     568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db     121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
```

```
Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              |||||||||||||||||||||||||||      ||||||   ||||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysArgLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              |||||||||||||||||||||    :::||||||:::|||:::|||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||   
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
              ||||||||||||||||||   ||||||||||||:::||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                 ||||||||||||:::::::::||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              :::||||||:::||||||||||||      ||||||||||||||||||||||||||||||
Db        481 IleAlaGluLeuGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              ||||||   |||||||||||||||||||||:::   |||   ||||||||||||||||||
Db        501 PheLysIleLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              |||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
                 |||||||||:::|||||||||||||||||||||||||||||||||||||||:::|||
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              ||||||||||||||||||:::|||||||||||||||||||||:::|||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::||||||||||||:::|||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::|||||||||||||||   |||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              |||
Db        721 Asp 721
```

RESULT 6
B44001
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (strain YU-2)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: B44001
R;Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.
J. Virol. 66, 6587-6600, 1992
A;Title: Complete nucleotide sequence, genome organization, and biological properties of
A;Reference number: A44001; MUID:93021387; PMID:1404605
A;Accession: B44001
A;Molecule type: DNA
A;Residues: 1-1003 <LIY>
A;Cross-references: UNIPROT:P35963; GB:M93258
C;Comment: This protein is synthesized as a gag-pol polyprotein.
C;Comment: Specific enzymatic cleavages may yield mature proteins including proteinase,
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;57-155/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status predicted

```
Alignment Scores:
Pred. No.:               9.68e-153      Length:         1003
Score:                   3450.00        Matches:        647
Percent Similarity:      94.87%         Conservative:   37
Best Local Similarity:   89.74%         Mismatches:     31
Query Match:             75.44%         Indels:         6
DB:                      1              Gaps:           3

US-09-610-313B-31 (1-2463) x B44001 (1-1003)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db      1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy    280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db     21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40

Qy    334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db     41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy    388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db     61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy    448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db     81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy    508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db    101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuHisGluAspPheArg 280

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyThrArgTyrGln 300

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db    321 ThrIleLeuGluProPheArgLysGlnAsnProAspLeuValIleTyrGlnTyrMetAsp 340

Qy   1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db    341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy   1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db    361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy   1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
Db    381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy   1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
Db    401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy   1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
Db    421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy   1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
Db    441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy   1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
Db    461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy   1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
Db    481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy   1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
Db    501 PheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520

Qy   1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
Db    521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy   1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
Db    541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy   1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
Db    561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy   1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
Db    581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600

Qy   2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
Db    601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnLysGlyArgGlnLys 620

Qy   2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
Db    621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy   2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
Db    641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy   2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
Db    661 IleGlnAlaGlnProAspArgSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
```

```
Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::|||||||||||||||   |||||||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              |||
Db        721 Asp 721

RESULT 7
GNVWA2
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate ARV-2)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C;Accession: A03968
R;Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Sh
Science 227, 484-492, 1985
A;Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).
A;Reference number: A04003; MUID:85090453; PMID:2578227
A;Accession: A03968
A;Molecule type: DNA
A;Residues: 1-1003 <SAN>
A;Cross-references: UNIPROT:P03369; GB:K02007; NID:g328658; PIDN:AAB59876.1; PID:g328662
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;57-147/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:
Pred. No.:              1.08e-152      Length:        1003
Score:                  3449.00        Matches:       646
Percent Similarity:     94.87%         Conservative:  38
Best Local Similarity:  89.60%         Mismatches:    31
Query Match:            75.42%         Indels:        6
DB:                     1              Gaps:          3

US-09-610-313B-31 (1-2463) x GNVWA2 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||   ||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------GACAACCCC 333
                 ||||||||||||||||||   ||||||||||||||   |||      :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||||:::||||||||||||:::      ||||||||||||||||||
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||||:::|||::::::||||||||:::||||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              |||||||||||||||||||||||||||||||||||||||||||||||||||||   :::
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProVal 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              ||||||||||||   ||||||||||||||||||||||:::||||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||||         ||||||   ||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              ||||||||||||||||||||     :::||||||:::||||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 ||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMet 400

Qy       1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
              ||||||||||||||||||   ||||||||||||:::||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                 ||||||||||||:::::::::||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              |||||||||||||||:::||||||||||||   |||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              ||||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        481 ValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500
```

```
Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              ||||||||||||||||||||||||||||||:::||||||   ||||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              |||||||||||||||||||||||||||||:::::: |||||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
                 ||||||||:::||||||||||||||||||||||||||   ||||||   :::|||
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::||||||:::   :::||||||||||||||||||||||||||||||   ||||||
Db        621 ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::||||||||||||||||   ||||||||||||||||||||||:::||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720

Qy       2362 GAT 2364
              |||
Db        721 Asp 721
```

RESULT 8
GNVWVL
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LV)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C;Accession: A03967
R;Muesing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.
Nature 313, 450-458, 1985
A;Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi
A;Reference number: A93355; MUID:85111157; PMID:2982104
A;Accession: A03967
A;Molecule type: DNA
A;Residues: 1-1012 <MUE>
A;Cross-references: UNIPROT:P03368
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;66-164/Product: retropepsin #status predicted <RTP>
F;90/Active site: Asp (shared with dimeric partner) #status experimental

Alignment Scores:
Pred. No.:   3.14e-152   Length: 1012

```
Score:                    3439.00       Matches:        647
Percent Similarity:       93.67%        Conservative:   34
Best Local Similarity:    89.00%        Mismatches:     28
Query Match:              75.20%        Indels:         18
DB:                       1             Gaps:           4

US-09-610-313B-31 (1-2463) x GNVWVL (1-1012)

Qy        238 GCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------------------------ 267
              |||||   |||||||||||||||||||
Db          4 AlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerPro 23

Qy        268 ------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG--- 318
                    ||||||||   ||||||||||||||||   ||||||||||||||
Db         24 ThrIleSerSerGluGlnThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrp 43

Qy        319 ---CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------ 369
                 |||   :::|||    ||||||||||||||:::||||||||||||:::
Db         44 GlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPhe 63

Qy        370 AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAG 429
              |||||||||||||||||||||||||||||||||||||||:::||||||:::||||||||||
Db         64 AsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGln 83

Qy        430 ATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTG 489
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 LeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeu 103

Qy        490 CCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAG 549
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 ProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGln 123

Qy        550 TACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGC 609
              ||||||||||||||||||||||||||||   |||||||||||||||||||:::|||
Db        124 TyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIleGlyThrValLeuValGly 143

Qy        610 CCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAAC 669
              |||||||||||||||||||||||||||||||:::||||||||:::|||||||||||||||
Db        144 ProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsn 163

Qy        670 TTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCC 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 PheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyPro 183

Qy        730 AAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAG 789
              |||||||||||||||||||||||||||||||||||||||||||||||      ||||||
Db        184 LysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThr 203

Qy        790 GAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCC 849
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        204 GluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrPro 223

Qy        850 GTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAG 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 ValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGlu 243

Qy        910 CTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGC 969
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 LeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGly 263

Qy        970 CTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCC 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 LeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValPro 283

Qy       1030 CTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACC 1089
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        284 LeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThr 303

Qy       1090 CCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATC 1149
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        304 ProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIle 323

Qy       1150 TTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTG 1209
```

```
Db        324 PheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleVal 343

Qy       1210 ATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGC 1263
Db        344 IleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArg 363

Qy       1264 GCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAG 1323
Db        364 ThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLys 383

Qy       1324 AAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGG 1383
Db        384 LysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrp 403

Qy       1384 ACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAG 1443
Db        404 ThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLys 423

Qy       1444 CTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTG 1503
Db        424 LeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeu 443

Qy       1504 TGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCC 1563
Db        444 CysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAla 463

Qy       1564 GAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTAC 1623
Db        464 GluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyr 483

Qy       1624 GACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTAC 1683
Db        484 AspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyr 503

Qy       1684 CAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACC 1743
Db        504 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 523

Qy       1744 GCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGC 1803
Db        524 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSer 543

Qy       1804 ATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAG 1863
Db        544 IleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGluThrTrpGlu 563

Qy       1864 ACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACC 1923
Db        564 ThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThr 583

Qy       1924 CCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACC 1983
Db        584 ProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThr 603

Qy       1984 TTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACC 2043
Db        604 PheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGlyLysAlaGlyTyrValThr 623

Qy       2044 GACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTG 2103
Db        624 AsnLysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeu 643

Qy       2104 CAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGC 2163
Db        644 GlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSer 663

Qy       2164 CAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAAC 2223
Db        664 GlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsn 683

Qy       2224 CAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCAC 2283
Db        684 GlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHis 703

Qy       2284 AAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2343
Db        704 LysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysIle 723

Qy       2344 CTGTTCCTGGACGGCATCGAT 2364
Db        724 LeuPheLeuAspGlyIleAsp 730
```

RESULT 9
GNLJSI
HIV-1 retropepsin (EC 3.4.23.16) - simian immunodeficiency virus SIVcpz
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: simian immunodeficiency virus SIVcpz
A;Note: host Pan troglodytes (chimpanzee)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: S09984
R;Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.
Nature 345, 356-359, 1990
A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
A;Reference number: S09983; MUID:90259077; PMID:2188136
A;Accession: S09984
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1027 <HUB>
A;Cross-references: UNIPROT:P17283; EMBL:X52154
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;81-180/Product: retropepsin #status predicted <RTP>
F;105/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:

Pred. No.:	1.17e-141	Length:	1027
Score:	3211.00	Matches:	595
Percent Similarity:	88.92%	Conservative:	63
Best Local Similarity:	80.41%	Mismatches:	77
Query Match:	70.22%	Indels:	5
DB:	1	Gaps:	2

US-09-610-313B-31 (1-2463) x GNLJSI (1-1027)

```
Qy        159 CTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAA 218
Db          7 LeuLeuAlaValTrpAlaArgGlyThrProAsnGluArgLeuHisArgLysThrGlyGlu 26

Qy        219 CTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCA 278
Db         27 -PhePheArgGluArgLeuAlaPheProGlnArgGluAlaArgGlnLeuCysAlaGluGl 46

Qy        279 GAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------GACAACCC 332
Db         46 nAsnArgThrAsnGlyProThrAspArgGluLeuTrpValProGlyGlyArgGluGluPr 66

Qy        333 CCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTG 392
Db         66 oGlyGluGluArgGlyArgGluGlnSerIleSerThrAsnLeuProGlnIleThrLeuTr 86

Qy        393 GCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACAC 452
Db         86 pGlnArgProLeuIleProValLysValGluGlyGlnLeuCysGluAlaLeuLeuAspTh 106

Qy        453 CGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGAT 512
Db        106 rGlyAlaAspAspThrValIleGluArgIleGlnLeuGlnGlyLeuTrpLysProLysMe 126

Qy        513 GATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGAT 572
Db        126 tIleGlyGlyIleGlyGlyPheIleLysValLysGlnPheAspAsnValHisIleGluIl 146
```

```
Qy    573 CTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGG 632
Db    146 eGluGlyArgLysValValGlyThrValLeuValGlyProThrProValAsnIleIleGl 166

Qy    633 CCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGAC 692
Db    166 yArgAsnIleLeuThrGlnLeuGlyCysThrLeuValPheProIleSerSerIleGluTh 186

Qy    693 CGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGAC 752
Db    186 rValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuSe 206

Qy    753 CGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGAT 812
Db    206 rAlaGluLysIleLysAlaLeuThrGluIleCysGlnGluMetGluLysGluGlyLysIl 226

Qy    813 CACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGA 872
Db    226 eSerLysIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAs 246

Qy    873 CAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTT 932
Db    246 pSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPh 266

Qy    933 CTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGAC 992
Db    266 eTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValTh 286

Qy    993 CGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTA 1052
Db    286 rValLeuAspValGlyAspAlaTyrPheSerCysProLeuAspLysAspPheArgLysTy 306

Qy   1053 CACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAA 1112
Db    306 rThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyValArgTyrGlnTyrAs 326

Qy   1113 CGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGAT 1172
Db    326 nValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSerMetThrLysIl 346

Qy   1173 CCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCT 1226
Db    346 eLeuGluProPheArgGluLysAsnProAspIleThrIleTyrGlnTyrMetAspAspLe 366

Qy   1227 GTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAA 1286
Db    366 uTyrValGlySerAspLeuGluIleAspGlnHisArgLysLysValGluGluLeuArgGl 386

Qy   1287 GCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTT 1346
Db    386 nHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPh 406

Qy   1347 CCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCC 1406
Db    406 eLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuPr 426

Qy   1407 CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGC 1466
Db    426 oGluLysGluValTrpThrValAsnAspIleGlnLysLeuIleGlyLysLeuAsnTrpAl 446

Qy   1467 CAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAA 1526
Db    446 aSerGlnIleTyrProGlyIleLysIleLysGlnLeuCysLysLeuIleArgGlyThrLy 466

Qy   1527 GGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCG 1586
Db    466 sLysLeuThrAspValValProLeuThrProGluAlaGluLeuGluLeuAlaGluAsnAr 486

Qy   1587 CGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGC 1646
Db    486 gGluIleValSerThrProValHisGlyValTyrTyrAspProAspLysGluLeuIleAl 506

Qy   1647 CGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAA 1706
Db    506 aGluIleGlnLysGlnGlyAsnCysGlnTrpThrTyrGlnIlePheGlnGluProHisLy 526

Qy   1707 GAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCA 1766
Db    526 sAsnLeuLysThrGlyLysTyrAlaArgGlnArgSerAlaHisThrAsnAspIleArgGl 546

Qy   1767 GCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCC 1826
Db    546 nLeuAlaGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThrPr 566

Qy   1827 CAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCA 1886
Db    566 oLysPheArgLeuProValGlnLysGluSerTrpGluAlaTrpTrpAlaGluTyrTrpGl 586

Qy   1887 GGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTA 1946
Db    586 nAlaThrTrpIleProGluTrpGluPheIleAsnThrProProLeuValLysLeuTrpTy 606

Qy   1947 CCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAA 2006
Db    606 rSerLeuGluThrGluProIleProThrThrAspThrTyrTyrValAspGlyAlaAlaAs 626

Qy   2007 CCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGT 2066
Db    626 nArgGluThrLysThrGlyLysAlaGlyTyrValThrAspLysGlyLysGlnLysIleIl 646

Qy   2067 GAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCA 2126
Db    646 eSerLeuGluAsnThrThrAsnGlnGlnAlaGluLeuLysAlaLeuLeuLeuAlaLeuGl 666

Qy   2127 GGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCA 2186
Db    666 nAspSerAspGlnGlnValAsnIleValThrAspSerGlnTyrValLeuGlyIleIleGl 686

Qy   2187 GGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAA 2246
Db    686 nSerGlnProAspHisSerGluSerGluLeuValAsnGlnIleIleGluGluLeuIleLy 706

Qy   2247 GAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCA 2306
Db    706 sLysGluLysIleTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGl 726

Qy   2307 GATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db    726 nValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 745
```

RESULT 10
S33980
pol polyprotein - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C;Accession: S33980
R;Carlini, F.
submitted to the EMBL Data Library, November 1991
A;Reference number: S33979
A;Accession: S33980
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-912 <CAR>
A;Cross-references: UNIPROT:Q78241; EMBL:Z11530; NID:g60192; PIDN:CAA77622.1; PID:g60194
C;Superfamily: pol polyprotein

```
Alignment Scores:
Pred. No.:              1.61e-136      Length:        912
Score:                  3100.00        Matches:       574
Percent Similarity:     95.87%         Conservative:  30
Best Local Similarity:  91.11%         Mismatches:    24
Query Match:            67.79%         Indels:        2
DB:                     2              Gaps:          1
```

US-09-610-313B-31 (1-2463) x S33980 (1-912)

```
Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db          1 MetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 20

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db         21 ValLysGlnTyrAspGlnIleAlaIleGluIleCysGlyHisLysAlaIleGlyThrVal 40

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db         41 LeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCys 60

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db         61 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 80

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db         81 AspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuIleGlu 100

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db        101 IleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyr 120

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db        121 AsnThrProValPheAlaIleLysLysLysAspGlyThrLysTrpArgLysLeuValAsp 140

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db        141 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 160

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db        161 ProGlyGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 180

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db        181 SerValProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsn 200

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db        201 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 220

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db        221 ProAlaIlePheGlnAlaSerMetThrLysIleLeuGluProPheArgLysGlnAsnPro 240

Qy       1201 GAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
Db        241 GluIleIleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Db        261 GlnHisArgThrLysIleGluGluLeuArgArgHisLeuLeuArgTrpGlyPheThrThr 280

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
Db        281 ProAspLysLysHisGlnLysGluProProPheLeuTrpIleGlyTyrGluLeuHisPro 300

Qy       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
Db        301 AspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAsp 320

Qy       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
Db        321 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLysVal 340

Qy       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
Db        341 ArgGlnLeuCysArgLeuLeuArgGlyAlaLysAlaLeuThrGluValIleProLeuThr 360

Qy       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
Db        361 LysGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysThrProValHisGly 380

Qy       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
Db        381 ValTyrTyrAspProSerLysAspLeuValAlaGluIleGlnLysGlnGlyLeuGlyGln 400

Qy       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
Db        401 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 420

Qy       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
Db        421 MetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 440

Qy       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
Db        441 ThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGlu 460

Qy       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
Db        461 ThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 480

Qy       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
Db        481 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 500

Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
Db        501 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly 520

Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
Db        521 TyrValThrAspArgGlyArgGlnLysValIleSerLeuThrAspThrThrAsnGlnLys 540

Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
Db        541 ThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleVal 560

Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
Db        561 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 580

Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
Db        581 LeuValSerGlnIleIleGluHisLeuIleLysLysGluLysValTyrLeuAlaTrpVal 600

Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
Db        601 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIle 620

Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db        621 ArgLysValLeuPheLeuAspGlyIleAsp 630
```

RESULT 11
T01668
pol polyprotein - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T01668
R;Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.
Cell 46, 63-74, 1986
A;Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol
A;Reference number: Z14389; MUID:86245056; PMID:2424612
A;Accession: T01668
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-902 <ALI>
A;Cross-references: EMBL:K03456; NID:g60228; PIDN:CAA28012.1; PID:g60230
C;Superfamily: pol polyprotein

Alignment Scores:

```
Pred. No.:                3.05e-136    Length:        902
Score:                    3094.00      Matches:       574
Percent Similarity:       97.10%       Conservative:  28
Best Local Similarity:    92.58%       Mismatches:    16
Query Match:              67.66%       Indels:        2
DB:                       2            Gaps:          1

US-09-610-313B-31 (1-2463) x T01668 (1-902)

Qy    511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
Db      1 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 20

Qy    571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
Db     21 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 40

Qy    631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
Db     41 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 60

Qy    691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
Db     61 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 80

Qy    751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
Db     81 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 100

Qy    811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
Db    101 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 120

Qy    871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
Db    121 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 140

Qy    931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
Db    141 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 160

Qy    991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
Db    161 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 180

Qy   1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
Db    181 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 200

Qy   1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
Db    201 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 220

Qy   1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
Db    221 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 240

Qy   1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
Db    241 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 260

Qy   1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
Db    261 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 280

Qy   1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404
Db    281 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 300

Qy   1405 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1464
Db    301 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 320

Qy   1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1524
Db    321 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 340

Qy   1525 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1584
Db    341 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 360

Qy   1585 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1644
Db    361 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 380

Qy   1645 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1704
Db    381 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 400

Qy   1705 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1764
Db    401 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 420

Qy   1765 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1824
Db    421 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 440

Qy   1825 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884
Db    441 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 460

Qy   1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1944
Db    461 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 480

Qy   1945 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2004
Db    481 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 500

Qy   2005 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2064
Db    501 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 520

Qy   2065 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2124
Db    521 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 540

Qy   2125 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2184
Db    541 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 560

Qy   2185 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2244
Db    561 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 580

Qy   2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2304
Db    581 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 600

Qy   2305 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db    601 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 620
```

```
RESULT 12
B47175
reverse transcriptase, AZT-sensitive variant - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: B47175
R;Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993
A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the
A;Reference number: A47175; MUID:93126353; PMID:7678340
A;Accession: B47175
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: DNA
```

```
A;Residues: 1-559 <MOH>
A;Cross-references: UNIPROT:Q9PXX1
A;Note: sequence extracted from NCBI backbone (NCBIP:122099)
C;Superfamily: pol polyprotein

Alignment Scores:
Pred. No.:              2.69e-121     Length:         559
Score:                  2771.00       Matches:        514
Percent Similarity:     96.06%        Conservative:   23
Best Local Similarity:  91.95%        Mismatches:     20
Query Match:            60.59%        Indels:         2
DB:                     2             Gaps:           1

US-09-610-313B-31 (1-2463) x B47175 (1-559)

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 20

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||   ||||||   ||||||
Db         21 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGlu 40

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db         41 MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal 60

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 80

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 100

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 120

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 140

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        141 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 160

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              |||||||||||||||||||||||||||||||||   :::||||||:::||||||||||||
Db        161 GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleValIle 180

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||                  ||||||||||||||||||||||||||||||||||||
Db        181 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr 200

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        201 LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLys 220

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 240

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||||||||   ||||||||||||:::||||||||||||||||||||||||||||||
Db        241 ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 260

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 280

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              |||||||||||||||   ||||||||||||:::::::::|||||||||||||||||||||
Db        281 LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGlu 300

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        301 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 320

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              |||||||||||||||:::||||||||||||||||||         |||||||||||||||
Db        321 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGln 340

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||:::||||||      |||
Db        341 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla 360

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              |||||||||||||||||||||||||||||||||||||||         ||||||||||||
Db        361 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle 380

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              ||||||||||||||||||:::|||:::|||||||||||||||||||||||||||||||||
Db        381 ValIleTrpGlyLysThrProArgPheLysLeuProIleGlnLysGluThrTrpGluThr 400

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 TrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 420

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        421 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 440

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||:::
Db        441 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 460

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              :::||||||||||||:::|||      ||||||      |||||||||||||||||||||
Db        461 LysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeuGln 480

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||   ||||||||||||||||||||     ||||||||||||||||||||||||||
Db        481 AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln 500

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln 520

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        521 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys 540

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2343
              ||||||||||||||||||:::|||||||||||||||      |||||||||||||||
Db        541 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 559

RESULT 13
A47175
reverse transcriptase, AZT-resistant variant - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A47175
R;Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993
A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the
A;Reference number: A47175; MUID:93126353; PMID:7678340
A;Accession: A47175
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: nucleic acid
A;Residues: 1-559 <MOH>
A;Cross-references: UNIPROT:Q9PXX2
A;Note: sequence extracted from NCBI backbone (NCBIP:122100)
C;Superfamily: pol polyprotein

Alignment Scores:
```

```
Pred. No.:                2.28e-120     Length:         559
Score:                    2751.00       Matches:        510
Percent Similarity:       95.89%        Conservative:   26
Best Local Similarity:    91.23%        Mismatches:     21
Query Match:              60.16%        Indels:         2
DB:                       2             Gaps:           1

US-09-610-313B-31 (1-2463) x A47175 (1-559)

Qy     673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
Db       1 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 20

Qy     733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
Db      21 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGlu 40

Qy     793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db      41 MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal 60

Qy     853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db      61 PheAlaIleLysLysLysAspSerThrArgTrpArgLysLeuValAspPheArgGluLeu 80

Qy     913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db      81 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 100

Qy     973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db     101 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 120

Qy    1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db     121 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 140

Qy    1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db     141 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 160

Qy    1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db     161 GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspMetValIle 180

Qy    1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db     181 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr 200

Qy    1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db     201 LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 220

Qy    1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
Db     221 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 240

Qy    1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
Db     241 ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 260

Qy    1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
Db     261 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys 280

Qy    1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
Db     281 LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleGlnLeuThrGluGluAlaGlu 300

Qy    1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
Db     301 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 320

Qy    1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
Db     321 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGln 340

Qy    1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
Db     341 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla 360

Qy    1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
Db     361 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle 380

Qy    1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
Db     381 ValIleTrpGlyLysIleProArgPheLysLeuProIleGlnLysGluThrTrpGluAla 400

Qy    1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
Db     401 TrpTrpIleGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 420

Qy    1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
Db     421 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 440

Qy    1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
Db     441 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 460

Qy    2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
Db     461 LysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGln 480

Qy    2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
Db     481 AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln 500

Qy    2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
Db     501 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln 520

Qy    2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
Db     521 IleIleGluGluLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys 540

Qy    2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2343
Db     541 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 559
```

RESULT 14
S46347
pol polyprotein - simian immunodeficiency virus SIVagm (isolate SAB-1)
C;Species: simian immunodeficiency virus SIVagm
A;Variety: isolate SAB-1
C;Date: 25-Dec-1994 #sequence_revision 14-Feb-1997 #text_change 26-Aug-1999
C;Accession: S46347
R;Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.; Barre-Sinoussi, F.; Hirsch, V.M.; A
EMBO J. 13, 2935-2947, 1994
A;Title: Mosaic genome structure of simian immunodeficiency virus from West African greer
A;Reference number: S46335; MUID:94298785; PMID:8026477
A;Accession: S46347
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1039 <JIN>
A;Cross-references: EMBL:U04005; NID:g466229; PIDN:AAA21505.1; PID:g466231
A;Experimental source: isolate SAB-1; sabaeus monkey
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
A;Note: this reading frame extends between two stop codons and does not begin with a star
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein

Alignment Scores:
Pred. No.:   9.73e-102   Length: 1039

```
Score:                     2350.00        Matches:         440
Percent Similarity:        74.07%         Conservative:    120
Best Local Similarity:     58.20%         Mismatches:      152
Query Match:               51.39%         Indels:          44
DB:                        2              Gaps:            7

US-09-610-313B-31 (1-2463) x S46347 (1-1039)

Qy       220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC------AAGGCCCGCGAGTTCCCCAGC 273
Db         1 PhePheArg---------ValTrpProLeuGlyGlnArgGluThrGlnGluPheProSer 17

Qy       274 GAGCAGAACCGCGCCAAC---AGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------ 324
Db        18 AspLeuHisGlnThrAsnSerSerProAsnGlyThrGlyLeuGlnGlnAlaGlyGlyLys 37

Qy       325 ------------------GACAACCCCCGCAGCGAGGCC--------------------- 345
Db        38 LeuValCysArgGlnThrSerAspGlnArgThrArgAlaArgArgSerSerAsnSerPro 57

Qy       346 ---------------------------GGCGCCGAGCGC--------------------- 357
Db        58 ValLysAlaValCysCysSerGlyGluThrAlaGluThrAlaValAlaLysProLeuAla 77

Qy       358 ---------------CAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
Db        78 ThrThrGluProLeuArgGlyGlyLeuGlnLeuProGlnValSerLeuTrpArgArgPro 97

Qy       403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
Db        98 MetLysThrValTyrIleGluGlyGlnLysValThrAlaLeuLeuAspThrGlyAlaAsp 117

Qy       463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
Db       118 AspSerValIleGlnGlyIleGluLeuGlyAspAsnTrpLysProArgIleIleGlyGly 137

Qy       523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
Db       138 IleGlyGlyCysIleAsnValLysAlaTyrHisAsnGlnGluValLysIleGluAspLys 157

Qy       583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
Db       158 ThrCysLysAlaThrIleLeuValGlyGluThrProValAsnIleIleGlyArgAsnVal 177

Qy       643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
Db       178 LeuAlaGlnLeuGlyValThrLeuAsnLeuThrGlnArgGluIleGluProIleLysVal 197

Qy       703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
Db       198 HisLeuLysProGlyGlnAspGlyProArgIleArgGlnTrpProLeuSerLysGluLys 217

Qy       763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
Db       218 IleGluAlaLeuLysAlaIleCysGluAspLeuGluLysGlnGlyHisLeuGluArgIle 237

Qy       823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
Db       238 GlyProGluAsnProTyrAsnThrProValPheAlaIleArgLysLysAspLysThrGln 257

Qy       883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
Db       258 TrpArgIleLeuMetAspPheArgGlnLeuAsnLysSerThrGlnAspPheGlnGluVal 277

Qy       943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
Db       278 GlnLeuGlyIleProHisProAlaGlyLeuGlnGlnArgGluGlnIleThrValLeuAsp 297

Qy      1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
Db       298 IleGlyAspAlaTyrPheSerCysProLeuAspProAspPheGlnLysTyrThrAlaPhe 317

Qy      1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
Db       318 ThrIleProSerValAsnAsnArgGluProGlyIleArgTyrGlnTyrLysValLeuPro 337

Qy      1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
Db       338 GlnGlyTrpLysGlySerProThrIlePheGlnThrThrAlaAsnLysIleLeuGlnGlu 357

Qy      1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGC 1236
Db       358 PheArgGlnLysAsnProAspValAspIleTyrGlnTyrMetAspAspMetLeuIleAla 377

Qy      1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
Db       378 SerAspArgProLysAlaGluHisLeuValMetValGlnGlnLeuArgAspTyrLeuGlu 397

Qy      1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATG 1356
Db       398 ThrTrpGlyPheLysThrProGluLysLysPheGlnLysAspProProTyrLeuTrpMet 417

Qy      1357 GGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1416
Db       418 GlyTyrGluLeuTyrProLysLysTrpGlnLeuGlnGluIleThrLeuProGluArgGlu 437

Qy      1417 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1476
Db       438 GluTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 457

Qy      1477 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1536
Db       458 TyrThrGlyIleLysThrLysHisLeuCysArgLeuIleArgGlyAlaArgProLeuThr 477

Qy      1537 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1596
Db       478 GluIleValGlnTrpThrGluGluAlaGluLeuGluLeuGluGluAsnArgGlnIleLeu 497

Qy      1597 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1656
Db       498 ArgGlnLysGlnGlyGlnTyrTyrAspProAlaLeuProLeuArgAlaLysValLeu 517

Qy      1657 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1716
Db       518 LysLeuGlyAspGlyGlnTrpGlyTyrGlnIleTyrGlnProGluAsnLysIleLeuLys 537

Qy      1717 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1776
Db       538 ValGlyLysTyrAlaLysIleLysThrAlaHisThrAsnGluLeuArgMetLeuAlaGly 557

Qy      1777 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1836
Db       558 LeuValGlnLysIleGlyLysGluSerIleValIleTrpGlyGlnIleProIleMetGlu 577

Qy      1837 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1896
Db       578 LeuProValGluArgGluLeuTrpGluGlnTrpTrpSerAspTyrTrpGlnValThrTrp 597

Qy      1897 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1956
Db       598 IleProGluTrpGluMetValSerThrProGlnLeuIleArgLeuTrpTyrLysLeuVal 617

Qy      1957 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2016
Db       618 LysAspProIleProGlyGluAlaValTyrTyrValAspGlyAlaAlaAsnArgAsnSer 637

Qy      2017 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2076
Db       638 LysGluGlyLysAlaGlyTyrLeuThrAspArgGlyAspGlnLysValValAlaLeuGlu 657

Qy      2077 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2136
Db       658 AsnThrThrAsnGlnLysAlaGluLeuGluAlaIleLeuLeuAlaLeuArgAspSerGly 677

Qy      2137 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2196
```

```
Db        678 SerLysValAsnIleIleThrAspSerGlnTyrAlaMetGlyIleIleAlaGlyGluPro 697

Qy       2197 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2256
Db        698 ThrGluSerAspAsnAsnIleValGlnGlnIleIleGluGluLeuIleLysLysGluAla 717

Qy       2257 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2316
Db        718 ValTyrIleAlaTrpValProAlaHisLysGlyValGlyGlyAsnGluGluIleAspLys 737

Qy       2317 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db        738 LeuValSerGlnGlyIleArgGlnValLeuPheLeuAspArgIleGlu 753
```

RESULT 15
S53092
pol polyprotein - human immunodeficiency virus type 2
C;Species: human immunodeficiency virus type 2, HIV-2
C;Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C;Accession: S53092
R;Becker, M.; Zorr, B.; Becker, A.; Habermehl, K.O.
submitted to the EMBL Data Library, March 1995
A;Description: Molecular and phylogenetic characterisation of a Guinea-Bissau-derived hu
A;Reference number: S53091
A;Accession: S53092
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1055 <BEC>
A;Cross-references: UNIPROT:O73194; EMBL:Z48731
C;Superfamily: pol polyprotein
C;Keywords: polyprotein

Alignment Scores:

Pred. No.:	3.36e-95	Length:	1055
Score:	2209.00	Matches:	417
Percent Similarity:	70.60%	Conservative:	128
Best Local Similarity:	54.02%	Mismatches:	186
Query Match:	48.31%	Indels:	41
DB:	2	Gaps:	10

US-09-610-313B-31 (1-2463) x S53092 (1-1055)

```
Qy        156 GGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCAC---CGAGCGCCA 212
Db          3 GlyLeuLeuGluMetTrpGln------------AspArgThrTyrHisGlyLysValPro 18

Qy        213 GGCCAACTT--------------------------------------------CTTCCG 227
Db         19 ArgGlnThrGlyGlyPhePheArgAspTrpProLeuGlyLysGluAlaProGlnLeuPro 38

Qy        228 CGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAA------ 281
Db         39 ArgGlyProGlySerAlaGlyAlaAsnThrAsnSerThrProSerArgSerSerSerGly 58

Qy        282 CCGCGCCAACAGCCCCACCAGCCGCGA------GCTGCAGGTGCGCGGCGACAACCCCCG 335
Db         59 ProThrGlyGluIleTyrAlaAlaArgGluLysAlaGluGlyAlaGluArgGluThrIle 78

Qy        336 CAGCGAGGCCGGCGC------CGAGCGCCA------GGGCACCCTGAACTT--------- 374
Db         79 GlnArgGlyAspArgGlyLeuAlaAlaProArgAlaGlyLysAspThrMetGlnGlyAsp 98

Qy        375 ------------------CCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
Db         99 AsnArgGlyPheAlaAlaPro-GlnPheSerLeuTrpAsnArgProValValThrAlaHi 118

Qy        417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
Db        118 sIleGluGlyGlnProValGluValLeuLeuAspThrGlyAlaAspAspSerIleValAl 138

Qy        477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db        138 aGlyIleGluLeuGlySerAsnTyrSerProLysIleValGlyGlyIleGlyGlyPheIl 158

Qy        537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
Db        158 eAsnThrLysGluTyrLysAsnValGluIleGluValLeuGlyLysArgValArgAlaTh 178

Qy        597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
Db        178 rIleMetThrGlyAspThrProIleAsnIlePheGlyArgAsnIleLeuThrAlaLeuGl 198

Qy        657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
Db        198 yMetSerLeuAsnLeuProValAlaLysIleGluProIleLysIleMetLeuLysProGl 218

Qy        717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
Db        218 yLysAspGlyProLysLeuArgGlnTrpProLeuThrLysGluLysIleGluAlaLeuLy 238

Qy        777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db        238 sGluIleCysGluLysMetGluArgGluGlyGlnLeuGluGluAlaProProThrAsnPr 258

Qy        837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db        258 oTyrAsnThrProThrPheAlaIleArgLysLysAspLysAsnLysTrpArgMetLeuIl 278

Qy        897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db        278 eAspPheArgGluLeuAsnLysValThrGlnAspPheThrGluIleGlnLeuGlyIlePr 298

Qy        957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db        298 oHisProAlaGlyLeuAlaLysLysArgArgIleThrValLeuAspValGlyAspAlaTy 318

Qy       1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db        318 rPheSerIleProLeuHisGluAspPheArgGlnTyrThrAlaPheThrLeuProSerVa 338

Qy       1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db        338 lAsnAsnAlaGluProGlyLysArgTyrIleTyrLysValLeuProGlnGlyTrpLysGl 358

Qy       1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db        358 ySerProAlaIlePheGlnTyrThrMetArgGlnValLeuGluProPheArgLysAlaAs 378

Qy       1197 CCCCGAGATCGTGATC------TACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250
Db        378 nGlnAspValIleIleIleGlnTyrMetAspAspIleLeuIleAlaSerAspArgThrAs 398

Qy       1251 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1310
Db        398 pLeuGluHisAspArgValValLeuGlnLeuLysGluLeuLeuAsnSerLeuGlyPheSe 418

Qy       1311 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCA 1370
Db        418 rThrProAspGluLysPheGlnLysAspProProTyrArgTrpMetGlyTyrGluLeuTr 438

Qy       1371 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1430
Db        438 pProThrLysTrpLysLeuGlnLysIleGlnLeuProGlnLysGluValTrpThrValAs 458

Qy       1431 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1490
Db        458 nAspIleGlnLysLeuValGlyValLeuAsnTrpAlaAlaGlnIleTyrProGlyIleLy 478

Qy       1491 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1550
Db        478 sThrLysHisLeuCysArgLeuIleArgGlyLysMetThrLeuThrGluGluIleGlnTr 498

Qy       1551 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1610
Db        498 pThrGluLeuAlaGluAlaGluLeuGluGluAsnArgValIleLeuSerGlnGluGlnGl 518
```

```
Qy       1611 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1670
                |||  ||||||         |||:::|||  |||   :::||||||          ::
Db        518 uGlyHisTyrTyrGlnGluGluLysGluLeuGluAlaThrValGlnLysAspGlnAspAs 538

Qy       1671 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1730
              :|||||||||||:::::::::|||         |||  ||||||  ||||||||||||
Db        538 nGlnTrpThrTyrLysValHisGlnGlyGlu---LysThrLeuLysValGlyLysTyrAl 557

Qy       1731 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1790
              ||||::::::      |||||||||  |||:::  |||   :::  ||||||||||||
Db        557 aLysValLysAsnThrHisThrAsnGlyValArgLeuLeuAlaGlnValValGlnLysIl 577

Qy       1791 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1850
              |      |||::::::||||||||||||:::   |||||||||    ||||||::::::::
Db        577 eGlyLysGluAlaLeuValIleTrpGlyArgIleProLysPheHisLeuProValGluAr 597

Qy       1851 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1910
              :|||  ||||||  ||||||  ||||||||||||   ||||||||||||:::|||::
Db        597 gGluIleTrpGluGlnTrpTrpAspAspTyrTrpGlnValThrTrpIleProAspTrpAs 617

Qy       1911 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1970
              :||||||:::|||||||||||||||:::|||   :::   |||   |||:::||||||
Db        617 pPheValSerThrProProLeuValArgLeuValPheAsnLeuValLysAspProIlePr 637

Qy       1971 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2030
                ||||||||||||||||    ||||||:::   ||||||::::::|||   ||||||||
Db        637 oGlyAlaGluThrPheTyrThrAspGlySerCysAsnArgGlnSerLysGluGlyLysAl 657

Qy       2031 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2090
              ||||||||||||||||:::||||||   |||:::        |||    :::||||||||||
Db        657 aGlyTyrValThrAspLysGlyArgAspLysValLysValLeuGluGlnThrThrAsnGl 677

Qy       2091 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2150
              |:::   ||||||:::|||         :::||||||  |||||||||    :::    |||||
Db        677 nGlnAlaGluLeuGluAlaPheAlaMetAlaLeuThrAspSerGlyProLysAlaAsnIl 697

Qy       2151 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2210
              |:::   ||||||||||||    :::||||||:::        ||||||    :::||||||::
Db        697 eIleValAspSerGlnTyrValMetGlyIleValValGlyGlnProThrGluSerGluAs 717

Qy       2211 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2270
              :    :::||||||||||||||||||::::::||||||||||||    :::|||:::::::||
Db        717 nArgIleValAsnGlnIleIleGluGluMetIleLysLysGluAlaIleTyrValAlaTr 737

Qy       2271 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2330
              ||||||||||||||||||||||||||||||:::::::::|||   ||||||||||:::||
Db        737 pValProAlaHisLysGlyIleGlyGlyAsnGlnGluValAspHisLeuValSerGlnGl 757

Qy       2331 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              ||||||:::||||||||||||:::   |||:::
Db        757 yIleArgGlnValLeuPheLeuGluLysIleGlu 768
```

Search completed: June 2, 2005, 05:08:48
Job time : 101.833 secs

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  1, 2005, 20:20:55 ; Search time 208.667 Seconds
                                           (without alignments)
                                           12088.668 Million cell updates/sec

Title:          US-09-610-313B-31
Perfect score:  4573
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2463

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters:      3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155136_15147/app_query.fasta_1.7893
-DB=UniProt_03 -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_725_@runat_31052005_155136_15147 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      UniProt_03:*
                1:  uniprot_sprot:*
                2:  uniprot_trembl:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4140	90.5	1427	2	Q9WF92	Q9wf92 human immun
2	4136	90.4	1427	2	Q9WF89	Q9wf89 human immun
3	4132	90.4	1428	2	Q9WF86	Q9wf86 human immun
4	3990	87.3	1427	2	Q9WF62	Q9wf62 human immun
5	3983	87.1	1429	2	Q6X4P8	Q6x4p8 human immun
6	3955	86.5	1425	2	Q6X4Q6	Q6x4q6 human immun
7	3948	86.3	1427	2	Q9WF96	Q9wf96 human immun
8	3944	86.2	1430	2	Q6X4R4	Q6x4r4 human immun
9	3935.5	86.1	1427	2	Q9WF71	Q9wf71 human immun
10	3931	86.0	1427	2	Q9WFA0	Q9wfa0 human immun
11	3928.5	85.9	1427	2	Q9WF65	Q9wf65 human immun
12	3927	85.9	1437	2	Q9WF77	Q9wf77 human immun
13	3919.5	85.7	1427	2	Q9WF74	Q9wf74 human immun
14	3913.5	85.6	1427	2	Q8ADX5	Q8adx5 human immun
15	3905	85.4	1425	2	Q9WF79	Q9wf79 human immun
16	3902	85.3	1427	2	Q9WF98	Q9wf98 human immun
17	3893.5	85.1	1429	2	Q9WF84	Q9wf84 human immun
18	3884	84.9	1428	2	Q6X4P0	Q6x4p0 human immun
19	3874.5	84.7	1427	2	Q9WF68	Q9wf68 human immun
20	3838.5	83.9	1432	2	Q8ADZ9	Q8adz9 human immun
21	3835	83.9	1433	2	Q8ADE9	Q8ade9 human immun
22	3833	83.8	1434	2	Q8ADV1	Q8adv1 human immun
23	3831	83.8	1430	2	Q7SV36	Q7sv36 human immun
24	3827	83.7	1425	2	Q7SV20	Q7sv20 human immun
25	3821.5	83.6	1433	2	Q8ADN7	Q8adn7 human immun
26	3820.5	83.5	1433	2	Q8ADK5	Q8adk5 human immun
27	3818.5	83.5	1433	2	Q8ADS7	Q8ads7 human immun
28	3815.5	83.4	1433	2	Q8ADG5	Q8adg5 human immun
29	3812.5	83.4	1435	2	Q9DQ33	Q9dq33 human immun
30	3811.5	83.3	1433	2	Q8ADE1	Q8ade1 human immun
31	3806.5	83.2	1431	2	Q8ADR1	Q8adr1 human immun
32	3805.5	83.2	1433	2	Q8ADC5	Q8adc5 human immun
33	3804.5	83.2	1433	2	Q8AE32	Q8ae32 human immun
34	3803.5	83.2	1433	2	Q8ADM1	Q8adm1 human immun
35	3803.5	83.2	1433	2	Q8ADU3	Q8adu3 human immun
36	3802.5	83.2	1437	2	O41782	O41782 human immun
37	3801	83.1	1434	2	Q8Q7Z7	Q8q7z7 human immun
38	3801	83.1	1435	2	Q9WLJ3	Q9wlj3 human immun
39	3800.5	83.1	1433	2	Q8AD85	Q8ad85 human immun
40	3798.5	83.1	1433	2	Q7SPT9	Q7spt9 human immun
41	3794.5	83.0	1433	2	Q8ADW7	Q8adw7 human immun
42	3794.5	83.0	1436	2	Q8AD93	Q8ad93 human immun
43	3792.5	82.9	1433	2	Q8AE23	Q8ae23 human immun
44	3791.5	82.9	1433	2	Q8ADR9	Q8adr9 human immun
45	3789.5	82.9	1433	2	Q8ADH3	Q8adh3 human immun

ALIGNMENTS

```
RESULT 1
Q9WF92
ID   Q9WF92      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF92;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110975; AAD17135.1; -.
DR   HSSP; P04585; 1JLA.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
```

```
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW   Protease; RNA-directed DNA polymerase; Transferase.
SQ   SEQUENCE   1427 AA;  161059 MW;  F3B0C741A1EFB10A CRC64;

Alignment Scores:
Pred. No.:              4.06e-156     Length:         1427
Score:                  4140.00       Matches:        781
Percent Similarity:     99.49%        Conservative:   1
Best Local Similarity:  99.36%        Mismatches:     2
Query Match:            90.53%        Indels:         2
DB:                     2             Gaps:           1

US-09-610-313B-31 (1-2463) x Q9WF92 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              |||||           ||||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139

Qy       2347 TTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 2
Q9WF89
ID   Q9WF89      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF89;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110974; AAD17126.1; -.
DR   HSSP; P04585; 1JLA.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  161293 MW;  270F25DF67615C06 CRC64;

Alignment Scores:
Pred. No.:              5.85e-156      Length:         1427
Score:                  4136.00        Matches:        780
Percent Similarity:     99.36%         Conservative:   1
Best Local Similarity:  99.24%         Mismatches:     3
Query Match:            90.44%         Indels:         2
DB:                     2              Gaps:           1

US-09-610-313B-31 (1-2463) x Q9WF89 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
```

```
                :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379

Qy        73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy       133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy       193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439

Qy       253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459

Qy       313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479

Qy       373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy       433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519

Qy       493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy       553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559

Qy       613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy       673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy       733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy       793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy       853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy       913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy       973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy      1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy      1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739

Qy      1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759

Qy      1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
             ||||||      ||||||||||||||||||||||||||||||||||||||||||||||||
Db       760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy      1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy      1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy      1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy      1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy      1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
             |||||||||||||||     ||||||||||||||||||||||||||||||||||||||||
Db       860 LysLeuLeuArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy      1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899

Qy      1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy      1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy      1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

Qy      1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy      1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy      1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019

Qy      1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy      2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy      2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy      2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099
```

```
Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139

Qy       2347 TTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

RESULT 3

```
Q9WF86
ID   Q9WF86      PRELIMINARY;      PRT;  1428 AA.
AC   Q9WF86;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110973; AAD17119.1; -.
DR   HSSP; P04585; 1JLA.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1428 AA;  161328 MW;  2D61CD9FEABD2387 CRC64;

Alignment Scores:
Pred. No.:              8.42e-156      Length:        1428
Score:                  4132.00        Matches:       779
Percent Similarity:     99.49%         Conservative:  3
Best Local Similarity:  99.11%         Mismatches:    2
Query Match:            90.36%         Indels:        2
DB:                     2              Gaps:          1

US-09-610-313B-31 (1-2463) x Q9WF86 (1-1428)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetMetGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||         |||||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAspArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139

Qy       2347 TTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 4
Q9WF62
ID   Q9WF62      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF62;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110961; AAD17021.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
```

```
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  160992 MW;  AC2ACBF190C77DAD CRC64;

Alignment Scores:
Pred. No.:                3.47e-150     Length:         1427
Score:                    3990.00       Matches:        746
Percent Similarity:       97.33%        Conservative:   19
Best Local Similarity:    94.91%        Mismatches:     19
Query Match:              87.25%        Indels:         2
DB:                       2             Gaps:           1

US-09-610-313B-31 (1-2463) x Q9WF62 (1-1427)

Qy     13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
Db    360 LeuAlaGluAlaMetGlyGlnAlaAsnSerMetAsnIleMetMetGlnArgSerAsnPhe 379

Qy     73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
Db    380 LysAsnProLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy    133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
Db    400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy    193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
Db    420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy    253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
Db    440 LysAlaArgGluPheProSerGluGlnThrArgAlaIleSerProThrSerArgGluLeu 459

Qy    313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
Db    460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluGlyGlnGlyAlaLeuAsn 479

Qy    373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
Db    480 LeuProGlnIleThrLeuTrpGlnArgProLeuValThrIleArgValGlyGlyGlnIle 499

Qy    433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
Db    500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519

Qy    493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
Db    520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy    553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
Db    540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy    613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
Db    560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy    673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
Db    580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy    733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
Db    600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy    793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db    620 MetGluMetGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy    853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db    640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy    913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db    660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy    973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db    680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy   1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db    700 AspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy   1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db    720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy   1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db    740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaLeuAsnProGluIleValIle 759

Qy   1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db    760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy   1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db    780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799

Qy   1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
Db    800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819
```

```
Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              |||||||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        820 ValGlnProIleAsnLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              |||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIle 959

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||:::|||||||||:::|||||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluGluGluProIleValGlyAlaGluThrPhe 1019

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ArgGlyArgGlnArgIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGlyValAsnIleValThrAspSerGln 1079

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              |||||||||||   ||||||:::||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyLysGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2347 TTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

RESULT 5
Q6X4P8

```
ID   Q6X4P8     PRELIMINARY;     PRT;  1429 AA.
AC   Q6X4P8;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14709249; DOI=10.1089/088922203771881220;
RA   Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,
RA   McCutchan F.E.;
RT   "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT   in full genome analysis, is not sustained in subgenomic regions.";
RL   AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., McCutchan F.E.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maayan S., Zeira M.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AY255825; AAP76538.1; -.
DR   HSSP; P03367; 1A8G.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
```

```
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
FT   NON_TER      1      1
SQ   SEQUENCE   1429 AA;  161586 MW;  D0B9CD124DC64E19 CRC64;

Alignment Scores:
Pred. No.:              6.56e-150      Length:         1429
Score:                  3983.00        Matches:        743
Percent Similarity:     97.47%         Conservative:   27
Best Local Similarity:  94.05%         Mismatches:     14
Query Match:            87.10%         Indels:         6
DB:                     2              Gaps:           2

US-09-610-313B-31 (1-2463) x Q6X4P8 (1-1429)

Qy     13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
          :::||||||||||||||||||    :::|||||||||:::||||||:::|||||||||
Db    358 LeuAlaGluAlaMetSerGlnAlaAsnAsnAlaAsnIleMetMetGlnLysSerAsnPhe 377

Qy     73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
          ||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db    378 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 397

Qy    133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
          |||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db    398 AsnCysArgAlaProArgLysArgGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 417

Qy    193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
          ||||||||||||||||||||||||||||||||||||:::|||||||||    |||||
Db    418 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheGlnGlnGly 437

Qy    253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
          :::|||||||||||||||||||||     ||||||||||||||||||||||||||
Db    438 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 457

Qy    313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGC--------- 363
          |||||||||||||||||||    ||||||||||||    |||||||||||
Db    458 GlnValArgGlyAspAsnProIleSerGluAlaGlyAspGluArgGlnGlyAlaLeuGln 477

Qy    364 ---ACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
             |||||||||    ||||||||||||||||||||||||||||||:::||||||:::
Db    478 ArgThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIle 497

Qy    421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
          ||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db    498 GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGlu 517

Qy    481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
          ::::::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    518 IleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 537

Qy    541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    538 ValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrVal 557

Qy    601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
          |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    558 LeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCys 577

Qy    661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    578 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 597

Qy    721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
          ||||||||||||||||||||||||||||||:::|||||||||:::|||||||||||||||
Db    598 AspGlyProLysValLysGlnTrpProLeuThrLysGluLysIleGluAlaLeuThrAla 617

Qy    781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    618 IleCysGluGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyr 637

Qy    841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    638 AsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAsp 657

Qy    901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    658 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 677

Qy    961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 ProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 697

Qy   1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
          ||||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db    698 SerValProLeuAspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsn 717

Qy   1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    718 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 737

Qy   1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
          |||:::|||||||||||||||||||||||||||||||||||||||||||:::||||||
Db    738 ProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnPro 757

Qy   1201 GAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
          |||:::||||||||||||      ||||||||||||||||||||||||||||||
Db    758 GluLeuValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 777

Qy   1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
          |||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db    778 GlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThr 797

Qy   1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 ProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro 817

Qy   1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
          ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db    818 AspLysTrpThrValGlnProIleGlnLeuProGluLysGluSerTrpThrValAsnAsp 837

Qy   1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    838 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal 857

Qy   1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThr 877

Qy   1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
          ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db    878 GluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGly 897

Qy   1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
          ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db    898 ValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGln 917

Qy   1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    918 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 937

Qy   1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    938 MetArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 957

Qy   1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
          ||||||   |||||||||||||||||||    |||||||||||||||||||||||||||
Db    958 MetGluGlyIleValIleTrpGlyLysValProLysPheArgLeuProIleGlnLysGlu 977

Qy   1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
```

```
                |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        978 ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProAspTrpGluPhe 997

Qy       1915 GTGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
              |||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        998 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGly 1017

Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ValGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1037

Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 TyrValThrAspArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLys 1057

Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1077

Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 1097

Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
              |||||||||||||||||||||||||||:::||||||:::|||||||||||||||||||||
Db       1098 LeuValAsnGlnIleIleGluGlnLeuIleArgLysGluArgValTyrLeuSerTrpVal 1117

Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
              |||||||||||||||||||||||||||||||||:::|||||||||||||||   ||||||
Db       1118 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIle 1137

Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||||||||||||||
Db       1138 ArgLysValLeuPheLeuAspGlyIleAsp 1147
```

RESULT 6
Q6X4Q6

```
ID   Q6X4Q6      PRELIMINARY;      PRT;  1425 AA.
AC   Q6X4Q6;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14709249; DOI=10.1089/088922203771881220;
RA   Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,
RA   McCutchan F.E.;
RT   "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT   in full genome analysis, is not sustained in subgenomic regions.";
RL   AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., McCutchan F.E.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maayan S., Zeira M.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AY255824; AAP76529.1; -.
DR   HSSP; Q72498; 1K6Y.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   1425 AA;  161115 MW;  782315D4D5305E18 CRC64;

Alignment Scores:
Pred. No.:              8.4e-149      Length:          1425
Score:                  3955.00       Matches:         734
Percent Similarity:     97.20%        Conservative:    30
Best Local Similarity:  93.38%        Mismatches:      20
Query Match:            86.49%        Indels:          2
DB:                     2             Gaps:            1

US-09-610-313B-31 (1-2463) x Q6X4Q6 (1-1425)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||:::||||||   :::   :::|||:::|||||||||:::||||||
Db        358 LeuAlaGluAlaMetThrGlnAlaAsnAsnThrHisIleMetMetGlnArgAsnAsnPhe 377

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||||||||||||||:::||||||||||||||||||||||||||||||:::||||||
Db        378 LysGlyProLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 397

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 417

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
```

```
Db     418 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluIleLeuAlaPheProGlnGly 437

Qy     253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
Db     438 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrAsnGlyGluLeu 457

Qy     313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
Db     458 GlnValArgArgAsnAsnProArgSerLysThrGlyValGluArgGlnGlyThrLeuAsn 477

Qy     373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
Db     478 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 497

Qy     433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
Db     498 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 517

Qy     493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
Db     518 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 537

Qy     553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
Db     538 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 557

Qy     613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
Db     558 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 577

Qy     673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
Db     578 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 597

Qy     733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
Db     598 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 617

Qy     793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db     618 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 637

Qy     853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db     638 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 657

Qy     913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db     658 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 677

Qy     973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db     678 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 697

Qy    1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db     698 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 717

Qy    1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db     718 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 737

Qy    1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db     738 GlnSerSerMetThrArgIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 757

Qy    1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db     758 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 777

Qy    1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db     778 LysIleGluGluLeuArgGlyHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 797

Qy    1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
Db     798 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 817

Qy    1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
Db     818 ValGlnProIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 837

Qy    1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
Db     838 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 857

Qy    1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
Db     858 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 877

Qy    1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
Db     878 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 897

Qy    1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
Db     898 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyAspAspGlnTrpThrTyrGln 917

Qy    1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
Db     918 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 937

Qy    1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
Db     938 HisThrAsnAspValLysGlnLeuValGlyAlaValGlnLysIleAlaMetGluGlyIle 957

Qy    1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
Db     958 ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 977

Qy    1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
Db     978 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProAspTrpGluPheValAsnThrPro 997

Qy    1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
Db     998 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyValGluThrPhe 1017

Qy    1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
Db    1018 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1037

Qy    2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
Db    1038 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1057

Qy    2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
Db    1058 AlaIleGlnIleAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1077

Qy    2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
Db    1078 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1097

Qy    2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
Db    1098 IleIleGluGlnLeuIleLysLysGluArgIleTyrLeuThrTrpValProAlaHisLys 1117

Qy    2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
Db    1118 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAsnGlyIleArgLysValLeu 1137

Qy    2347 TTCCTGGACGGCATCGAT 2364
Db    1138 PheLeuAspGlyIleAsp 1143
```

```
RESULT 7
Q9WF96
ID   Q9WF96       PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF96;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110976; AAD17144.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  160518 MW;  FBDA73D603A68B18 CRC64;

Alignment Scores:
Pred. No.:              1.59e-148       Length:         1427
Score:                  3948.00         Matches:        730
Percent Similarity:     97.07%          Conservative:   33
Best Local Similarity:  92.88%          Mismatches:     21
Query Match:            86.33%          Indels:         2
DB:                     2               Gaps:           1

US-09-610-313B-31 (1-2463) x Q9WF96 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::|||||||||||||||||||||   :::   ||||||::::::|||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetIleGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::|||   :::|||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              |||||||||||||||||||||||||||||||||   |||:::||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheGlyGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              :::|||   |||||||||||||||||||||   ||||||:::   |||||||||   |||
Db        460 ArgValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||:::
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysValGlyGlyGlnVal 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||::::::||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              |||||||||   |||||||||||||||||||||||||||||||||||||||:::||||||
Db        540 AspGlnIleThrIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              |||||||||||||||||||||||||||:::|||||||||:::||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   ||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        700 AspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              |||||||||||||||||||||||||||||||||||:::||||||   |||||||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||            ||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||   ||||||||:::||||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||||||||:::||||||:::|||:::||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              ||||||||||||||||||||||||   |||||||||||||||||||   |||||||||||
Db        940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              |||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||||
Db       1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||   ||||||:::::||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              ||||||||||||||||||||:::||||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2347 TTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 8
Q6X4R4
ID   Q6X4R4      PRELIMINARY;      PRT;  1430 AA.
AC   Q6X4R4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14709249; DOI=10.1089/088922203771881220;
RA   Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,
RA   McCutchan F.E.;
RT   "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT   in full genome analysis, is not sustained in subgenomic regions.";
RL   AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., McCutchan F.E.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maayan S., Zeira M.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AY255823; AAP76521.1; -.
DR   HSSP; P12497; 1B9D.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
```

```
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   1430 AA;  161540 MW;  5BBA39B85D03FA0B CRC64;

Alignment Scores:
Pred. No.:              2.29e-148      Length:         1430
Score:                  3944.00        Matches:        731
Percent Similarity:     97.09%         Conservative:   36
Best Local Similarity:  92.53%         Mismatches:     17
Query Match:            86.25%         Indels:         6
DB:                     2              Gaps:           2

US-09-610-313B-31 (1-2463) x Q6X4R4 (1-1430)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||   |||:::   ||||||:::||||||:::||||||||
Db        359 LeuAlaGluAlaMetSerGlnValThrAsnThrAsnIleMetMetGlnLysSerAsnPhe 378

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||::::::   |||||||||||||||||||||||||||||||:::||||||
Db        379 LysGlyProArgLysThrIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 398

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              |||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db        399 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyArgGluGlyHisGlnMet 418

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||:::|||||||||     ||||||
Db        419 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheGlnGlnGly 438

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::|||||||||||||||   ||||||   ||||||||||||||||||||||||||||||
Db        439 GluAlaArgGluPheProProGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 458

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC------------CAG 360
              |||||||||||||||||||||   ||||||||||||    ||||||            |||
Db        459 GlnValArgGlyAspAsnProLeuSerGluAlaGlyThrGluArgProGlyThrLeuGln 478

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||   |||||||||||||||||||||||||||||||||:::||||||:::
Db        479 GlyThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIle 498

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGlu 518

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ::::::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        519 IleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 538

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        539 ValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlySerVal 558

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 LeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCys 578

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              |||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        579 ThrLeuAsnPheProIleSerProIleGluThrIleProValLysLeuLysProGlyMet 598

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 AspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAla 618

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              |||||||||:::||||||:::|||||||||||||:::|||||||||||||||||||||||
Db        619 IleCysGluAspMetGluArgGluGlyLysIleSerLysIleGlyProGluAsnProTyr 638

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        639 AsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAsp 658

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 678

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 ProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 698

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||:::|||||||||||||||||||||||||||||||||   |||
Db        699 SerValProLeuAspGluAsnPheArgLysTyrThrAlaPheThrIleProSerThrAsn 718

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 738

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              |||:::|||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        739 ProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnPro 758

Qy       1201 GAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              ||||||||||||||||||            ||||||||||||||||||||||||||||||
Db        759 GluIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 778

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              |||||||||||||||:::||||||||||||   ||||||||||||||||||||||||||
Db        779 GlnHisArgAlaLysValGluGluLeuArgGlyHisLeuLeuArgTrpGlyPheThrThr 798

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 ProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro 818
```

```
Qy       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
              ||||||||||||||||||    |||:::|||||||||||:::||||||||||||||||||
Db        819 AspLysTrpThrValGlnThrIleGlnLeuProGluLysAspSerTrpThrValAsnAsp 838

Qy       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal 858

Qy       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
              |||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        859 ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValAlaLeuThr 878

Qy       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        879 GluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGly 898

Qy       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        899 ValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGln 918

Qy       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        919 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 938

Qy       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        939 ArgArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 958

Qy       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 MetGluSerIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGlu 978

Qy       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        979 ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 998

Qy       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||
Db        999 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 1018

Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1038

Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
              ||||||||||||:::|||||||||||||:::|||||||||    ||||||||||||||||
Db       1039 TyrValThrAspLysGlyArgGlnLysValValSerLeuAsnGluThrThrAsnGlnLys 1058

Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1059 ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1078

Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||
Db       1079 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerAspSerGlu 1098

Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
              :::|||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db       1099 IleValAsnGlnIleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpVal 1118

Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
              |||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db       1119 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerLysGlyIle 1138

Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||||||||||||||
Db       1139 ArgLysValLeuPheLeuAspGlyIleAsp 1148
```

RESULT 9
Q9WF71

```
ID   Q9WF71      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF71;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110964; AAD17047.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  161326 MW;  B9FC43C29B26CFA2 CRC64;
```

Alignment Scores:

```
Pred. No.:                 4.96e-148     Length:          1427
Score:                     3935.50       Matches:         727
Percent Similarity:        96.95%        Conservative:    35
Best Local Similarity:     92.49%        Mismatches:      21
Query Match:               86.06%        Indels:          3
DB:                        2             Gaps:            2

US-09-610-313B-31 (1-2463) x Q9WF71 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||    |||:::   |||:::::::||||||||||||||||
Db        361 LeuAlaGluAlaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||||:::||||||:::||||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              |||||||||||||||||||||||||||         ||||||   ||||||      |||
Db        460 GlnValArgGlyAspAsnProArgSerGluThrArgAlaGluGlyGlnGlyAsnPheAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              :::||||||||||||||||||||||||||||||||||||||||||:::::::::||||||
Db        500 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysIleIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              :::||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        540 GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProIleAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   ||||||||||||||||||||||||||||||||||||||||||      ||||||
Db        700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||    :::||||||:::|||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||      ||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||:::|||||||||:::||||||    ||||||||||||||||||
Db        780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              |||||||||||:::||||||:::|||:::|||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              :::|||||||||||||||||||||||||||||||||||||:::|||:::    ||||||
Db        840 IleGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysHisLeuCys 859

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              ||||||||||||||:::||||||:::||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              ||||||||||||||||||||||||||||||      ||||||||||||||||||||   |||
Db        940 HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleAlaMetGluCysIle 959

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              |||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||||||||||||      ||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
```

```
                 :::||||||||||||||||||||||   ||||||||||||||||   |||||||||
Db       1040 LysGlyArgGlnLysIleValSerLeuAsnGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              |||||||||||||||||||:::|||||||||||||||   |||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2347 TTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 10
Q9WFA0
ID   Q9WFA0      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WFA0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110978; AAD17162.1; -.
DR   PIR; S49086; S49086.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  160503 MW;  24C410797550C0F8 CRC64;

Alignment Scores:
Pred. No.:               7.47e-148     Length:         1427
Score:                   3931.00       Matches:        727
Percent Similarity:      96.95%        Conservative:   35
Best Local Similarity:   92.49%        Mismatches:     22
Query Match:             85.96%        Indels:         2
DB:                      2             Gaps:           1

US-09-610-313B-31 (1-2463) x Q9WFA0 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||||   :::   ||||||:::::::||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetIleGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||||:::|||     :::||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||:::|||||||||||:::||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAspPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::|||||||||||||||||||||||   |||||||||||||||||||||    ||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerGlyGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||   |||||||||||||||||||||    ||||||:::   |||||||||   |||
Db        460 GlnValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnVal 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||:::::||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539
```

```
Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              |||||||||   |||||||||||||||||||||||||||||||||||||||:::||||||
Db        540 AspGlnIleThrIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||:::|||||||||:::|||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   |||||||||||||||||||||||||||||||||||||||   |||||||||
Db        700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||:::||||||   ||||||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||            ||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              |||||||||||||||   |||||||||:::||||||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              |||||||||:::||||||:::|||:::|||||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              |||||||||||||||||||||   |||||||||||||||   ||||||||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValLeuTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProLysTrpGluPheValAsnThrPro 999

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||:::|||||||||||||   |||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyCysValThrAsp 1039

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db       1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||   ||||||::::::|||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              ||||||||||||:::||||||:::|||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyHisGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2347 TTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 11
Q9WF65
ID   Q9WF65      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF65;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
```

```
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110962; AAD17030.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW   Protease; RNA-directed DNA polymerase; Transferase.
SQ   SEQUENCE   1427 AA;  161347 MW;  595F19865DE81D63 CRC64;

Alignment Scores:
Pred. No.:              9.38e-148      Length:         1427
Score:                  3928.50        Matches:        728
Percent Similarity:     96.95%         Conservative:   34
Best Local Similarity:  92.62%         Mismatches:     21
Query Match:            85.91%         Indels:         3
DB:                     2              Gaps:           2

US-09-610-313B-31 (1-2463) x Q9WF65 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||   |||:::   |||::::::||||||||||||||||||
Db        361 LeuAlaGluAlaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::||||||:::||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::|||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||||||      ||||||   |||||||||||||||
Db        460 GlnValArgGlyAspAsnProArgSerGluThrArgAlaGluGlyGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              :::||||||||||||||||||||||||||||||||||||||||||:::::::::||||||
Db        500 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              :::||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        540 GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   |||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||   ||||||||||||||||||||   :::||||||:::|||||||||
Db        740 GlnSerSerMetIleLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||            ||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779
```

```
Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||:::|||||||||:::||||||    |||||||||   ||||||
Db        780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProTyrLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||||||||:::||||||:::|||:::||||||||||||:::|||||||||||||||
Db        820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValHisAspIleGlnLysLeu 839

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||:::|||:::   ||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysHisLeuCys 859

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              |||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              |||||||||||||||:::||||||:::|||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              |||||||||||||||||||||||||||||||   ||||||||||||   ||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleThrMetGluSerIle 959

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              ||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              :::|||||||||||||||   |||||||||||||||||||||:::   |||||||||
Db       1040 LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnArgAlaGluLeuGln 1059

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              ||||||||||||||||||||:::||||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2347 TTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 12
Q9WF77
ID   Q9WF77      PRELIMINARY;      PRT;  1437 AA.
AC   Q9WF77;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110967; AAD17072.1; -.
DR   HSSP; P24736; 1NCP.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
```

```
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1437 AA;  161978 MW;  65DAA2562FC317A6 CRC64;

Alignment Scores:
Pred. No.:              1.08e-147       Length:         1437
Score:                  3927.00         Matches:        736
Percent Similarity:     96.23%          Conservative:   30
Best Local Similarity:  92.46%          Mismatches:     18
Query Match:            85.87%          Indels:         12
DB:                     2               Gaps:           3

US-09-610-313B-31 (1-2463) x Q9WF77 (1-1437)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::|||||||||||||||||||||   |||   ||||||:::||||||:::|||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaAsnSerValAsnIleMetMetGlnLysSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::|||   :::||||||||||||||||||||||||||||||||||||:::
Db        380 LysGlyProArgArgAsnValLysCysPheAsnCysGlyLysGluGlyHisIleAlaLys 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCC------------------AACAGC 294
              :::|||||||||||||||   ||||||   ||||||                  ||||||
Db        440 GluAlaArgGluPheProProGluGlnIleArgAlaSerSerProAsnSerThrAsnSer 459

Qy        295 CCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAG 354
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 ProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGlu 479

Qy        355 ------------CGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
                             |||||||||||||||   |||||||||||||||||||||||||||
Db        480 GlyGlnGlyThrLeuGlnGlyThrLeuAsnCysProGlnIleThrLeuTrpGlnArgPro 499

Qy        403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 LeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAlaAsp 519

Qy        463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
              :::|||||||||||||||::::::||||||||||||||||||||||||||||||||||||
Db        520 AsnThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 539

Qy        523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        540 IleGlyGlyPheIleLysValArgGlnTyrAspGlnIleValIleGluIleCysGlyLys 559

Qy        583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        560 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 579

Qy        643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 599

Qy        703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        600 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpThrLeuThrGluGluLys 619

Qy        763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        620 IleLysAlaLeuThrGluIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 639

Qy        823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 GlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLys 659

Qy        883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 679

Qy        943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp 699

Qy       1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
              :::|||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        700 MetGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 719

Qy       1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 739

Qy       1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        740 GlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluPro 759

Qy       1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGC 1236
              ||||||   :::||||||||||||||||||||||||            ||||||||||||
Db        760 PheArgLeuGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 779

Qy       1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
              |||||||||||||||||||||   ||||||:::|||||||||||||||:::|||||||||
Db        780 SerAspLeuGluIleGlyGlnArgArgAlaGlnIleGluGluLeuArgGluHisLeuLeu 799

Qy       1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATG 1356
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 819

Qy       1357 GGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1416
              ||||||||||||||||||||||||||||||||||||||||||:::||||||:::|||:::
Db        820 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProAspLysAsp 839

Qy       1417 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1476
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 859

Qy       1477 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThr 879

Qy       1537 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1596
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 AspValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu 899

Qy       1597 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1656
              :::|||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        900 LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln 919

Qy       1657 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 LysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 939

Qy       1717 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 ThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThrGlu 959

Qy       1777 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1836
              ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db        960 AlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThrProLysPheArg 979

Qy       1837 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 LeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrp 999
```

```
Qy       1897 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 1019

Qy       1957 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2016
              ||||||||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db       1020 LysGluProIleProGlyValGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 1039

Qy       2017 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2076
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 LysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSerLeuThr 1059

Qy       2077 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGly 1079

Qy       2137 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 SerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 1099

Qy       2197 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2256
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db       1100 AspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleGlnLysGluTrp 1119

Qy       2257 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2316
              |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db       1120 ValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLys 1139

Qy       2317 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              |||||||||:::||||||||||||:::||||||||||||||||||||||
Db       1140 LeuValSerGlnGlyIleArgLysMetLeuPheLeuAspGlyIleAsp 1155
```

RESULT 13
Q9WF74

```
ID   Q9WF74      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF74;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110965; AAD17056.1; -.
DR   PIR; S49086; S49086.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  161299 MW;  924682CF5095D763 CRC64;
```

Alignment Scores:

Pred. No.:	2.13e-147	Length:	1427
Score:	3919.50	Matches:	724
Percent Similarity:	96.82%	Conservative:	37
Best Local Similarity:	92.11%	Mismatches:	22
Query Match:	85.71%	Indels:	3
DB:	2	Gaps:	2

US-09-610-313B-31 (1-2463) x Q9WF74 (1-1427)

```
Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||   :::   :::::::::||||||:::|||||||||
Db        361 LeuAlaGluAlaMetSerGlnAla---AsnThrSerValMetMetGlnLysSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::||||||:::|||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              |||||||||||||||||||||||||||         |||   ||||||      |||
Db        460 GlnValArgGlyAspAsnProArgSerGluThrArgValGluGlyGlnGlyAsnPheAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy         433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
               :::||||||||||||||||||||||||||||||||||||||||||:::::::::||||||
Db         500 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519

Qy         493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy         553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
               :::||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db         540 GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy         613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
               |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnPheGlyCysThrLeuAsnPhe 579

Qy         673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy         733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db         600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp 619

Qy         793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy         853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy         913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy         973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy        1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
               ||||||   ||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719

Qy        1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
               |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db         720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy        1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
               ||||||||||||||||||||||||||||||||||||   :::||||||:::|||||||||
Db         740 GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759

Qy        1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
               ||||||                  ||||||||||||||||||||||||||||||||||||||||||
Db         760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy        1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
               ||||||||||||||||||:::|||||||||:::||||||   ||||||||||||||||||
Db         780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys 799

Qy        1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy        1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
               ||||||||||||:::||||||:::|||:::||||||||||||||||||||||||||||||
Db         820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy        1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
               ||||||||||||||||||||||||||||||||||||||||||:::||||||   ||||||
Db         840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValArgHisLeuCys 859

Qy        1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy        1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
               ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db         880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy        1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
               |||||||||||||||:::||||||:::|||||||||||||||||||||||||||||||||
Db         900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy        1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy        1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
               ||||||||||||||||||||||||||||||   |||:::|||||||||||||||||||||
Db         940 HisThrAsnAspValLysGlnLeuThrGluValValArgLysIleAlaMetGluSerIle 959

Qy        1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
               |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db         960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979

Qy        1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy        1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
               ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019

Qy        1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
               |||||||||||||||||||||||||||||||||:::|||||||||||||||||||||:::
Db        1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 1039

Qy        2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
               :::||||||||||||||||||   ||||||||||||||||||||||||   |||||||||
Db        1040 LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy        2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy        2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099

Qy        2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
               |||||||||||||||||||||||||||::::::|||||||||||||||||||||||||||
Db        1100 IleIleGluGlnLeuIleLysLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119

Qy        2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
               |||||||||||||||||||||:::|||||||||||||||   ||||||||||||||||||
Db        1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy        2347 TTCCTGGACGGCATCGAT 2364
               ||||||||||||||||||
Db        1140 PheLeuAspGlyIleAsp 1145

RESULT 14
Q8ADX5
ID   Q8ADX5      PRELIMINARY;      PRT;  1427 AA.
AC   Q8ADX5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
GN   Name=gag-pol;
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22375625; PubMed=12487816; DOI=10.1089/088922202320886325;
RA   Harris M.E., Serwadda D., Sewankambo N., Wabwire F., Kim B.,
RA   Kigozi G., Kiwanuka N., Phillips J.B., Meehen M., Lutalo T.,
RA   Lane J.R., Merling R., Gray R., Wawer M., Birx D.L., Robb M.L.,
RA   McCutchan F.E.;
RT   "Among 46 near full length HIV type 1 genome sequences from Rakai
RT   District, Uganda, subtype D and AD recombinants predominate.";
RL   AIDS Res. Hum. Retroviruses 18:1281-1290(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L., Robb M.L.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., Phillips J.B., Lane J.R., Merling R., McCutchan F.E.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Lutalo T.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Meehen M., Wawer M.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Serwadda S., Sewankambo N., Wabwire F., Kigozi G., Kiwanuka N.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF484491; AAN73556.1; -.
DR   HSSP; P24736; 1NCP.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW   Protease; RNA-directed DNA polymerase; Transferase.
FT   NON_TER       1      1
SQ   SEQUENCE   1427 AA;  161058 MW;  20559041E2A88B5C CRC64;

Alignment Scores:
Pred. No.:              3.67e-147     Length:        1427
Score:                  3913.50       Matches:       724
Percent Similarity:     96.69%        Conservative:  36
Best Local Similarity:  92.11%        Mismatches:    23
Query Match:            85.58%        Indels:        3
DB:                     2             Gaps:          2

US-09-610-313B-31 (1-2463) x Q8ADX5 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||        :::||||||:::||||||:::        |||
Db        361 LeuAlaGluAlaMetSerGlnValGlnHisSerAsnIleMetMetGlnGlnGlyLysPhe 380

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              :::|||   ||||||   ||||||||||||||||||||||||||||||:::||||||
Db        381 ArgGlyGlnLysArg---IleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::|||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||   |||||||||||||||   |||||||||||||||
Db        460 GlnValArgGlyAspAsnProArgPheGluAlaGlyAlaGluGlyGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||   |||:::|||:::||||||||||||:::
Db        480 PheProGlnIleThrLeuTrpGlnArgProPheValAlaIleArgValGlyGlyGlnLeu 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||:::   ||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleLysLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyAsnTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp 619
```

```
Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIle 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db        680 LysLysArgLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db        700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db        740 GlnAsnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuAspIleGluGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db        780 LysIleGluGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
Db        820 ValGlnProIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys 859

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
Db        900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGln 919

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysValArgThrAla 939

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
Db        940 HisThrAsnAspIleLysGlnLeuThrGluAlaValGlnLysIleAlaArgGluSerIle 959

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
Db        960 IleIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluAla 979

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleProGlyValGluThrPhe 1019

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysMetGlyLysAlaGlyTyrValThrAsp 1039

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
Db       1040 ArgGlyArgGlnLysIleIleSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnValValThrAspSerGln 1079

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnLys 1099

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2347 TTCCTGGACGGCATCGAT 2364
Db       1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 15
Q9WF79
ID   Q9WF79      PRELIMINARY;      PRT;  1425 AA.
AC   Q9WF79;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110969; AAD17087.1; -.
DR   HSSP; P24736; 1NCP.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
```

```
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1425 AA;  160569 MW;  6DC8A1AEF3C88DA7 CRC64;

Alignment Scores:
Pred. No.:              7.97e-147     Length:         1425
Score:                  3905.00       Matches:        727
Percent Similarity:     96.82%        Conservative:   34
Best Local Similarity:  92.49%        Mismatches:     21
Query Match:            85.39%        Indels:         4
DB:                     2             Gaps:           2

US-09-610-313B-31 (1-2463) x Q9WF79 (1-1425)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||||   ::::::||||||:::||||||::::::||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaAsnAsnSerAsnIleMetMetGlnLysAsnAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||   ||||||:::||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProThrArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysLysAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||:::||||||||||||||||||:::|||||||||||||||||
Db        420 LysAspCysThrGluArgArgAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||         ||||||||||||||||||:::|||
Db        440 GluAlaArgGluPheProSerGluGlnThrGlyAlaAsnSerProThrSerArgLysLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
                    ||||||:::||||||   ||||||||||||   ||||||||||||||||||
Db        460 ------ArgGlyAsnAsnProCysSerGluAlaGlyAspGluArgGlnGlyThrLeuAsn 477

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        478 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 497

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              |||||||||||||||||||||||||||||||||||||||||||||||||::::::|||||
Db        498 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 517

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 GlyLysTrpArgProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 537

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              |||||||||   |||||||||||||||||||||||||||||||||||||:::||||||
Db        538 AspGlnIleProIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 557

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        558 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPhe 577

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 597

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 617

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 637

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 657

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 677

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 697

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 717

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 737

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||:::|||||||||||:::|||||
Db        738 GlnSerSerMetThrLysIleLeuGluProPheArgAlaLysAsnProGluLeuValIle 757

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||      ||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 777

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 ArgIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 797

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 817

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||   |||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        818 ValGlnSerIleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 837

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
```

```
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        838 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyValLysValArgGlnLeuCys 857

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        858 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProProThrGluGluAlaGlu 877

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        878 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 897

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              |||||||||||||||:::|||||||||||||||||||||   ||||||||||||||||||
Db        898 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlyAspGlnTrpThrTyrGln 917

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 937

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        938 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleSerMetGluSerIle 957

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              |||||||||||||||   |||||||||||||||||||||||||||||||||   ||||||
Db        958 ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluAlaTrpGluAla 977

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 997

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||   ||||||:::   |||||||||||||||
Db        998 ProLeuValLysLeuTrpTyrGlnLeuGluThrGluProMetAlaGlyAlaGluThrPhe 1017

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1037

Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
              :::|||||||||::::::|||:::||||||||||||||||||||||||   |||||||||
Db       1038 LysGlyArgGlnGluValValThrLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1057

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
              ||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db       1058 AlaIleGlnLeuAlaLeuGlnAspSerGlyProGluValAsnIleValThrAspSerGln 1077

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1097

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1117

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
              ||||||||||||||||||::::::|||||||||||||||   |||||||||:::||||||
Db       1118 GlyIleGlyGlyAsnGluLysValAspLysLeuValSerSerGlyIleArgGluValLeu 1137

Qy       2347 TTCCTGGACGGCATCGAT 2364
              ||||||||||||||||||
Db       1138 PheLeuAspGlyIleAsp 1143
```

Search completed: June 2, 2005, 05:04:08
Job time : 290.667 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  2, 2005, 04:23:54 ; Search time 153.167 Seconds
                                          (without alignments)
                                          11117.427 Million cell updates/sec

Title:          US-09-610-313B-31
Perfect score:  4573
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2463

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1465611 seqs, 345679903 residues

Total number of hits satisfying chosen parameters:      2931222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155138_15193/app_query.fasta_1.7893
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09610313_@CGN_1_1_550_@runat_31052005_155138_15193
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_AA:*
                1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
                17:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
                19:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                20:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                 SUMMARIES
                %
Result          Query
```

No.	Score	Match	Length	DB	ID	Description
1	3570	78.1	998	16	US-10-332-413-4	Sequence 4, Appli
2	3523.5	77.1	850	16	US-10-332-413-10	Sequence 10, Appl
3	3497	76.5	1003	14	US-10-283-847-17	Sequence 17, Appl
4	3475	76.0	1015	17	US-10-634-165-9	Sequence 9, Appli
5	3452	75.5	739	16	US-10-093-953A-26	Sequence 26, Appl
6	3452	75.5	739	16	US-10-093-953A-31	Sequence 31, Appl
7	3430	75.0	1003	16	US-10-325-468-23	Sequence 23, Appl
8	3428	75.0	1003	16	US-10-325-468-35	Sequence 35, Appl
9	3427	74.9	1003	16	US-10-325-468-9	Sequence 9, Appli
10	3427	74.9	1003	16	US-10-325-468-20	Sequence 20, Appl
11	3419	74.8	995	15	US-10-296-734-2	Sequence 2, Appli
12	3405	74.5	995	15	US-10-296-734-1470	Sequence 1470, Ap
13	3251.5	71.1	1006	15	US-10-296-734-1471	Sequence 1471, Ap
14	3183.5	69.6	1014	14	US-10-301-661A-6	Sequence 6, Appli
15	3119.5	68.2	1350	10	US-09-952-060-35	Sequence 35, Appl
16	3119.5	68.2	1350	16	US-10-380-641-35	Sequence 35, Appl
17	3119.5	68.2	1350	17	US-10-636-730-35	Sequence 35, Appl
18	2822.5	61.7	875	10	US-09-952-060-6	Sequence 6, Appli
19	2822.5	61.7	875	15	US-10-168-217A-6	Sequence 6, Appli
20	2822.5	61.7	875	16	US-10-380-641-6	Sequence 6, Appli
21	2822.5	61.7	875	17	US-10-636-730-6	Sequence 6, Appli
22	2817	61.6	850	10	US-09-952-060-2	Sequence 2, Appli
23	2817	61.6	850	15	US-10-168-217A-2	Sequence 2, Appli
24	2817	61.6	850	16	US-10-380-641-2	Sequence 2, Appli
25	2817	61.6	850	17	US-10-636-730-2	Sequence 2, Appli
26	2798.5	61.2	875	10	US-09-952-060-8	Sequence 8, Appli
27	2798.5	61.2	875	15	US-10-168-217A-8	Sequence 8, Appli
28	2798.5	61.2	875	16	US-10-380-641-8	Sequence 8, Appli
29	2798.5	61.2	875	17	US-10-636-730-8	Sequence 8, Appli
30	2793	61.1	850	10	US-09-952-060-4	Sequence 4, Appli
31	2793	61.1	850	15	US-10-168-217A-4	Sequence 4, Appli
32	2793	61.1	850	16	US-10-380-641-4	Sequence 4, Appli
33	2793	61.1	850	17	US-10-636-730-4	Sequence 4, Appli
34	2789	61.0	562	9	US-09-735-487-14	Sequence 14, Appl
35	2789	61.0	562	14	US-10-342-188-14	Sequence 14, Appl
36	2785	60.9	560	15	US-10-399-920-2	Sequence 2, Appli
37	2783	60.9	561	14	US-10-059-271-87	Sequence 87, Appl
38	2777	60.7	562	14	US-10-102-622-10	Sequence 10, Appl
39	2758	60.3	560	14	US-10-205-641-1	Sequence 1, Appli
40	2748.5	60.1	561	14	US-10-102-622-12	Sequence 12, Appl
41	2736	59.8	560	11	US-09-725-652-1	Sequence 1, Appli
42	2702	59.1	546	8	US-08-808-031A-29	Sequence 29, Appl
43	2689	58.8	979	14	US-10-271-181B-118	Sequence 118, App
44	2588	56.6	851	14	US-10-059-271-88	Sequence 88, Appl
45	2571	56.2	999	15	US-10-346-000A-3	Sequence 3, Appli

```
                                ALIGNMENTS

RESULT 1
US-10-332-413-4
; Sequence 4, Application US/10332413
; Publication No. US20040116660A1
; GENERAL INFORMATION:
;  APPLICANT: Johnstone, Robert Edward
;  APPLICANT:  Swanstrom, Ronald Ivar
;  APPLICANT:  Morris, Lynn
;  APPLICANT:  Karim, Salim Abdool
;  APPLICANT:  Williamson, Carolyn
;  TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H
;  TITLE OF INVENTION:  Subtype Isolates, Their Genes nd Modifications and Derivatives T
;  FILE REFERENCE: 45669-281993
;  CURRENT APPLICATION NUMBER: US/10/332,413
;  CURRENT FILING DATE:  2003-09-12
;  PRIOR APPLICATION NUMBER: PCT/IB01/01208
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/216,995
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: ZA 2000/3437
;  PRIOR FILING DATE: 2000-07-10
;  PRIOR APPLICATION NUMBER: ZA 2000/4924
```

```
;  PRIOR FILING DATE: 2000-09-15
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 998
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-10-332-413-4

Alignment Scores:
Pred. No.:              1.56e-165    Length:          998
Score:                  3570.00      Matches:         666
Percent Similarity:     97.07%       Conservative:    29
Best Local Similarity:  93.02%       Mismatches:      19
Query Match:            78.07%       Indels:          2
DB:                     16           Gaps:            1

US-09-610-313B-31 (1-2463) x US-10-332-413-4 (1-998)

Qy      223 TTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAAC 282
Db        1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20

Qy      283 CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAG 342
Db       21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu 40

Qy      343 GCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
Db       41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro 60

Qy      403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
Db       61 LeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp 80

Qy      463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
Db       81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 100

Qy      523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
Db      101 IleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLys 120

Qy      583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
Db      121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140

Qy      643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
Db      141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160

Qy      703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
Db      161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180

Qy      763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
Db      181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200

Qy      823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
Db      201 GlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220

Qy      883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
Db      221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240

Qy      943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
Db      241 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp 260

Qy     1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
Db      261 ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 280

Qy     1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
Db      281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 300

Qy     1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
Db      301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320

Qy     1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGC 1236
Db      321 PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 340

Qy     1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
Db      341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeu 360

Qy     1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATG 1356
Db      361 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380

Qy     1357 GGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1416
Db      381 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400

Qy     1417 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1476
Db      401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420

Qy     1477 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1536
Db      421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440

Qy     1537 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1596
Db      441 AspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu 460

Qy     1597 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1656
Db      461 LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln 480

Qy     1657 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1716
Db      481 LysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 500

Qy     1717 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1776
Db      501 ThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThrGlu 520

Qy     1777 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1836
Db      521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540

Qy     1837 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1896
Db      541 LeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThrTrp 560

Qy     1897 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1956
Db      561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 580

Qy     1957 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2016
Db      581 LysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 600

Qy     2017 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2076
Db      601 LysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeuSer 620

Qy     2077 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2136
Db      621 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640
```

```
Qy       2137 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2196
              |||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        641 SerGluValAsnIleIleThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 660

Qy       2197 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2256
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        661 AspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluArg 680

Qy       2257 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2316
              |||||||||||||||||||||||||   |||:::  |||||||||   :::||||||
Db        681 ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700

Qy       2317 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
              ||||||||:::||||||||||||||||||||   ||||||||||||||
Db        701 LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716
```

RESULT 2
US-10-332-413-10
; Sequence 10, Application US/10332413
; Publication No. US20040116660A1
; GENERAL INFORMATION:
; APPLICANT: Johnstone, Robert Edward
; APPLICANT: Swanstrom, Ronald Ivar
; APPLICANT: Morris, Lynn
; APPLICANT: Karim, Salim Abdool
; APPLICANT: Williamson, Carolyn
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H
; TITLE OF INVENTION: Subtype Isolates, Their Genes nd Modifications and Derivatives T
; FILE REFERENCE: 45669-281993
; CURRENT APPLICATION NUMBER: US/10/332,413
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: PCT/IB01/01208
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/216,995
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: ZA 2000/3437
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: ZA 2000/4924
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 850
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-332-413-10

```
Alignment Scores:
Pred. No.:              2.8e-163     Length:         850
Score:                  3523.50      Matches:        668
Percent Similarity:     91.05%       Conservative:   34
Best Local Similarity:  86.64%       Mismatches:     44
Query Match:            77.05%       Indels:         25
DB:                     16           Gaps:           5
```

US-09-610-313B-31 (1-2463) x US-10-332-413-10 (1-850)

```
Qy         97 TGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAG 156
                 |||      ||||||                          |||       |||
Db         84 CysAlaThrValGlyLys------------GlyAspArgCysGlyProLeuArgTyrTyr 99

Qy        157 GGCTGCTGGAAG------------------------TGCGGCAAGGAGGGCCACCAGATG 192
                       |||:::                         |||::::::|||
Db        100 AlaSerTrpArgLysGlyAspValLeuGlnGlyAspValGlyArgGlnGlyPheProSer 119

Qy        193 AAGGACTGCACCGAGCGCCAG------------GCCAACTTCTTCCGCGAGGACCTG--- 237
                 |||       :::|||:::             :::       |||      |||
Db        120 HisAspValValLysArgArgProValProSerLeuHisAlaCysArgSerThrLeuGlu 139

Qy        238 ------------------GCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
                                :::|||||||||||   ||||||:::||||||||||||||
Db        140 AspProArgValProSerSerPheProGlnGlyProAlaArgGlnPheProSerGluGln 159

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
                 |||||||||||||||||||||||||||||||||||||||      ||||||||||||||
Db        160 ThrArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAspAsnProArgSer 179

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGC 399
              |||     |||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        180 GluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg 199

Qy        400 CCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCC 459
              ||||||||||||||||:::|||||||||    :::|||||||||||||||||||||||||
Db        200 ProLeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAla 219

Qy        460 GACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGC 519
              ||||||||||||||||||||:::::::::|||||||||||||||||||||||||||||||
Db        220 AspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGly 239

Qy        520 GGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGC 579
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGly 259

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        260 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 279

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 299

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluVal 319

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 LysIleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLys 339

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
              |||||||||||||||||||||||||||:::||||||||||||||:::|||||||||
Db        340 IleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysGluAspSerThr 359

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 379

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 399

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
              |||||||||||||||||||||||||||||||||||||   |||||||||||||||||||
Db        400 AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAla 419

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 439

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
              ||||||||||||||||||||||:::|||||||||:::|||||||||||||||||||||||
Db        440 ProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSerMetThrLysIleLeuGlu 459

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTG 1233
              |||||||||||:::|||||||||||||||||||||||||      |||   |||||||||
Db        460 ProPheArgAlaLysAsnProGluIleValIleTyrGlnTyrMetAlaAlaLeuTyrVal 479

Qy       1234 GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTG 1293
              |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        480 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeu 499

Qy       1294 CTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGG 1353
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 519
```

```
Qy   1354 ATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAG 1413
Db    520 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLys 539

Qy   1414 GAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG 1473
Db    540 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpThrSerGln 559

Qy   1474 ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTG 1533
Db    560 IleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 579

Qy   1534 ACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATC 1593
Db    580 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIle 599

Qy   1594 CTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATC 1653
Db    600 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 619

Qy   1654 CAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTG 1713
Db    620 GlnLysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 639

Qy   1714 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACC 1773
Db    640 LysThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThr 659

Qy   1774 GAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTC 1833
Db    660 GluAlaValGlnLysIleSerLeuGluSerIleValThrTrpGlyLysThrProLysPhe 679

Qy   1834 CGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACC 1893
Db    680 ArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThr 699

Qy   1894 TGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTG 1953
Db    700 TrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeu 719

Qy   1954 GAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAG 2013
Db    720 GluLysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 739

Qy   2014 ACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTG 2073
Db    740 ThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeu 759

Qy   2074 ACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGC 2133
Db    760 SerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 779

Qy   2134 GGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAG 2193
Db    780 GluSerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln 799

Qy   2194 CCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAG 2253
Db    800 ProAspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlu 819

Qy   2254 AAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGAC 2313
Db    820 ArgAlaTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAspGluGlnValAsp 839

Qy   2314 AAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
Db    840 LysLeuValSerSerGlyIleArgLysValLeu 850
```

RESULT 3
US-10-283-847-17
; Sequence 17, Application US/10283847
; Publication No. US20030162720A1
; GENERAL INFORMATION:
; APPLICANT: Nicolette, Charles A.
; APPLICANT: Walker, Bruce
; TITLE OF INVENTION: THERAPEUTIC ANTI-HIV (IV9) COMPOUNDS
; FILE REFERENCE: GZ 2112.00
; CURRENT APPLICATION NUMBER: US/10/283,847
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 60/345,116
; PRIOR FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus
US-10-283-847-17

Alignment Scores:

Pred. No.:	5.48e-162	Length:	1003
Score:	3497.00	Matches:	655
Percent Similarity:	95.56%	Conservative:	34
Best Local Similarity:	90.85%	Mismatches:	26
Query Match:	76.47%	Indels:	6
DB:	14	Gaps:	3

US-09-610-313B-31 (1-2463) x US-10-283-847-17 (1-1003)

```
Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy    280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db     21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy    334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db     41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60

Qy    388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db     61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy    448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db     81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy    508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db    101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
```

```
Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              ||||||||||||||||||      :::||||||:::||||||||||||||      
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 |||||||||||||||||||||||||||||||||||||||     ||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::||||||||||||||||||    |||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                 |||||||||||::::::::||||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              :::|||||||||||||||||||||      |||||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              |||||||||||||||||||||||||||||||:::||||||    ||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              ||||||||||||||||||||||||||||||        ||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              |||||||||||||||||||:::|||||||||||||||||||||:::||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::|||:::||||||:::||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              ||||||||||||||      ||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::||||||||||||||||    |||||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              |||
Db        721 Asp 721
```

RESULT 4
US-10-634-165-9
; Sequence 9, Application US/10634165
; Publication No. US20050095581A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Nancy T
; APPLICANT: Gallo, Robert C.
; APPLICANT: Wong-Staal, Flossie
; TITLE OF INVENTION: DETECTION OF HIV-1 DNA
; FILE REFERENCE: 223695
; CURRENT APPLICATION NUMBER: US/10/634,165
; CURRENT FILING DATE: 2003-08-05
; PRIOR APPLICATION NUMBER: 08/463,028
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 06/693,866
; PRIOR FILING DATE: 1985-01-23
; PRIOR APPLICATION NUMBER: 06/659,339
; PRIOR FILING DATE: 1984-10-10
; PRIOR APPLICATION NUMBER: 06/643,306
; PRIOR FILING DATE: 1984-08-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
;   LENGTH: 1015
;   TYPE: PRT
;   ORGANISM: T cell leukemia-lymphoma virus (HTLV)
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1015)
;   OTHER INFORMATION: /note= "pol protein of HTLV-III"
US-10-634-165-9

```
Alignment Scores:
Pred. No.:               6.42e-161     Length:        1015
Score:                   3475.00       Matches:       654
Percent Similarity:      93.72%        Conservative:  33
Best Local Similarity:   89.22%        Mismatches:    28
Query Match:             75.99%        Indels:        18
DB:                      17            Gaps:          4

US-09-610-313B-31 (1-2463) x US-10-634-165-9 (1-1015)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
              ||||||||||||||||||||||    |||||||||||||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
```

```
Qy    268 -----------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db     21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy    304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db     41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy    358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db     61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy    412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db     81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy    472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db    101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy    532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db    121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy    592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db    141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy    652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db    161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy    712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db    181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy    772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db    201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy    832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db    221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy    892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db    241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy    952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db    261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy   1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db    281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy   1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db    301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy   1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db    321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy   1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db    341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy   1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db    361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy   1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
Db    381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy   1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
Db    401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy   1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
Db    421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy   1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
Db    441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy   1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
Db    461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy   1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
Db    481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy   1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
Db    501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy   1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
Db    521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy   1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
Db    541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy   1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
Db    561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy   1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
Db    581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy   1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
Db    601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy   2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
Db    621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy   2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
Db    641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy   2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
Db    661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy   2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
Db    681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy   2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
Db    701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy   2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db    721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 5

```
US-10-093-953A-26
; Sequence 26, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
;  APPLICANT: Robinson, Harriet L.
;  APPLICANT:  Smith, James M.
;  APPLICANT:  Hua, Jian
;  APPLICANT:  Moss, Bernard
;  APPLICANT:  Amara, Rama
;  APPLICANT:  Wyatt, Linda
;  APPLICANT:  Earl, Patricia
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
;  TITLE OF INVENTION:  AN IMMUNE RESPONSE
;  FILE REFERENCE: 12804-005002
;  CURRENT APPLICATION NUMBER: US/10/093,953A
;  CURRENT FILING DATE:  2002-03-08
;  PRIOR APPLICATION NUMBER: US 60/186,364
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 60/251,083
;  PRIOR FILING DATE: 2000-12-01
;  PRIOR APPLICATION NUMBER: US 09/798,675
;  PRIOR FILING DATE: 2001-03-02
;  PRIOR APPLICATION NUMBER: PCT/US01/06795
;  PRIOR FILING DATE: 2001-03-02
;  PRIOR APPLICATION NUMBER: US 60/324,845
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/325,004
;  PRIOR FILING DATE: 2001-09-26
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
;   LENGTH: 739
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: protein encoded by construct of vaccine vector
;   OTHER INFORMATION: pGA2 and insert JS2 expressing clade HIV-1 VL
US-10-093-953A-26

Alignment Scores:
Pred. No.:              8.24e-160      Length:        739
Score:                  3452.00        Matches:       651
Percent Similarity:     93.45%         Conservative:  34
Best Local Similarity:  88.81%         Mismatches:    30
Query Match:            75.49%         Indels:        18
DB:                     16             Gaps:          4

US-09-610-313B-31 (1-2463) x US-10-093-953A-26 (1-739)

Qy     220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db       1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy     280 AACCGC------------------------------------GCCAACAGCCCCACCAGC 303
Db      21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40

Qy     304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db      41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy     358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db      61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy     412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db      81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy     472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db     101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy     532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db     121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy     592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db     141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy     652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db     161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy     712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db     181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy     772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db     201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy     832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db     221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy     892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db     241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy     952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db     261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy    1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db     281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy    1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db     301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy    1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db     321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy    1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db     341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360

Qy    1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db     361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy    1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
Db     381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy    1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
Db     401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy    1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
Db     421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440

Qy    1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
Db     441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy    1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
Db     461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy    1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
```

```
               ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
                        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
              |||||:::||||||   ||||||||||||||||||   ||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540

Qy       1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
              ||||||     ||||||||||||||||||||||||||||||||||:::||||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
              |||:::||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
              ||||||||||||||||:::::||||||||||||:::|||   ||||||   ||||||
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
              ||||||||||||:::||||||||||||   ||||||||||||||||||||   ||||||
Db        641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
              :::|||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
                 ||||||||||:::||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 6
US-10-093-953A-31
; Sequence 31, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; APPLICANT: Moss, Bernard
; APPLICANT: Amara, Rama
; APPLICANT: Wyatt, Linda
; APPLICANT: Earl, Patricia
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; TITLE OF INVENTION: AN IMMUNE RESPONSE
; FILE REFERENCE: 12804-005002
; CURRENT APPLICATION NUMBER: US/10/093,953A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/798,675
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/325,004
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 739
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: protein encoded by construct of vaccine vector
;   OTHER INFORMATION: pGA1 and vaccine insert expressing clade B HIV-1
;   OTHER INFORMATION: Gag-Pol
US-10-093-953A-31

```
Alignment Scores:
Pred. No.:              8.24e-160    Length:         739
Score:                  3452.00      Matches:        651
Percent Similarity:     93.45%       Conservative:   34
Best Local Similarity:  88.81%       Mismatches:     30
Query Match:            75.49%       Indels:         18
DB:                     16           Gaps:           4

US-09-610-313B-31 (1-2463) x US-10-093-953A-31 (1-739)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||   ||||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGC----------------------------------GCCAACAGCCCCACCAGC 303
                 |||                                  |||||||||||||||
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
              |||||||||||||||      |||   :::|||   |||||||||||||||:::|||
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
              |||||||||:::      ||||||||||||||||||||||||||||||||||||:::
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
              ||||||||||||:::|||||||||||||||||||||||||||||||||:::|||||||||
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
              |||      ||||||   ||||||||||||||||||||||||:::|||||||||||||||
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
```

```
Qy    832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db    221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy    892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db    241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy    952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db    261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy   1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db    281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy   1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db    301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy   1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db    321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy   1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db    341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360

Qy   1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db    361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy   1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAG 1365
Db    381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy   1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1425
Db    401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy   1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1485
Db    421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440

Qy   1486 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1545
Db    441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy   1546 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605
Db    461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy   1606 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1665
Db    481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy   1666 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
Db    501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy   1726 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
Db    521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540

Qy   1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1845
Db    541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy   1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905
Db    561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy   1906 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1965
Db    581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy   1966 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2025
Db    601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy   2026 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2085
Db    621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy   2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2145
Db    641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy   2146 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2205
Db    661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy   2206 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2265
Db    681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy   2266 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325
Db    701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy   2326 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
Db    721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 7
US-10-325-468-23
; Sequence 23, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.10 protein Pol
US-10-325-468-23

Alignment Scores:

Pred. No.:	9.85e-159	Length:	1003
Score:	3430.00	Matches:	638
Percent Similarity:	94.73%	Conservative:	45
Best Local Similarity:	88.49%	Mismatches:	32
Query Match:	75.01%	Indels:	6
DB:	16	Gaps:	3

US-09-610-313B-31 (1-2463) x US-10-325-468-23 (1-1003)

```
Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db      1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy    280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db     21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
```

```
Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCAGATCACC 387
              ||||||||||||||:::||||||||||||:::      |||||||||||||||||
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              ||||||||||||||||||||:::||||||:::|||||||||:::|||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||:::|||   ||||||||:::||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              :::||||||||   |||||||:::|||||||||||||:::||||||||||||||||||||
Db        121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||||:::||||||||||||||:::||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              |||||||||||||||||||||||||          ||||||    |||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||::::::||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              |||   ||||||||||||||      :::||||||:::|||:::||||||||
Db        321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 |||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::|||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
              ||||||||||||||||||    ||||||||||||:::||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                 ||||||||||::::::::||||||||:::||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              :::    ||||||||||||||||||      |||||||||||||||||:::|||||||||
Db        481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              ||||||||||||||||||||||||||||||||||      |||:::||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgSerAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              |||||||||||||||||||||||||||||||||      ||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
                 ||||||||:::||||||||||||||||||||||||||||||||||||:::|||
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              ||||||||||||||||:::||||||||||||||||||||||      ||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::||||||||||||:::||||||||||||||||||||||||||||||||    ||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              ||||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::||||||||||||||||    ||||||||||:::||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              :::
Db        721 Glu 721
```

RESULT 8
US-10-325-468-35
; Sequence 35, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:

```
;  APPLICANT: Soong, Nay Wei
;  APPLICANT:  Pekrun, Katja
;  APPLICANT:  Shibata, Riri
;  TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
;  TITLE OF INVENTION:  MODELS OF HIV-1 PATHOGENESIS
;  FILE REFERENCE: 0166.210US
;  CURRENT APPLICATION NUMBER: US/10/325,468
;  CURRENT FILING DATE:  2002-12-19
;  PRIOR APPLICATION NUMBER: US 60/343,524
;  PRIOR FILING DATE: 2001-12-21
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone P8A26 protein Pol
US-10-325-468-35

Alignment Scores:
Pred. No.:              1.23e-158      Length:         1003
Score:                  3428.00        Matches:        639
Percent Similarity:     94.73%         Conservative:   44
Best Local Similarity:  88.63%         Mismatches:     32
Query Match:            74.96%         Indels:         6
DB:                     16             Gaps:           4

US-09-610-313B-31 (1-2463) x US-10-325-468-35 (1-1003)

Qy     220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
           ||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db       1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy     280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGAC---AACCCC 333
              ||||||||||||||      ||||||    ||||||   |||||||||    |||
Db      21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgGlyAspAsnAsnSer 40

Qy     334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
              |||||||||||||:::||||||||||||:::       ||||||||||||||||
Db      41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy     388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
           ||||||||||||||||||||:::||||||:::||||||||||:::|||||||||||||
Db      61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy     448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
           ||||||||||||||||||||||||||||:::|||    |||||||||:::||||||||
Db      81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy     508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
           ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db     101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy     568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
           :::|||||||||    ||||||:::|||||||||||||:::||||||||||||||||||
Db     121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy     628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
           |||||||||||:::||||||||:::||||||||||||||||||||||||||||||||||
Db     141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy     688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy     748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
           |||||||||||||||||||||||||||         ||||||   ||||||||||||||||
Db     181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy     808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
           ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy     868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
           ||||||||||||||||||||||||||||||||||||||||||||||||::::::||||||
Db     221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy     928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy     988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db     261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy    1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy    1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
           ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db     301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy    1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
           |||   ||||||||||||||    :::||||||:::|||:::||||||||
Db     321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy    1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                ||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db     341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy    1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
           |||:::||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db     361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy    1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy    1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
           ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db     401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy    1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
           |||||||||||||||||    ||||||||||||:::||||||||||||||||||||||||
Db     421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy    1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
           ||||||||||||:::::::::|||||||||||||||||||||||||||||||||||||
Db     441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy    1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
           ||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db     461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy    1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
           :::     ||||||||||||||||           ||||||||||||||||:::|||||||||
Db     481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy    1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
           |||||||||||||||||||||||||||||||   |||    |||||||||||||||||
Db     501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520

Qy    1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
           |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db     521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy    1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
                |||||||:::|||||||||||||||||||||||||||||||||||||:::|||
Db     541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy    1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
```

```
Qy     1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
Db      581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy     2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
Db      601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy     2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
Db      621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy     2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
Db      641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy     2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
Db      661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy     2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
Db      681 IleLysLysGluLysValTyrLeuThrTrpIleProAlaHisLysGlyIleGlyGlyAsn 700

Qy     2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
Db      701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy     2362 GAT 2364
            :::
Db      721 Glu 721
```

RESULT 9
US-10-325-468-9
; Sequence 9, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone 1.4 protein Pol;
;   OTHER INFORMATION: clone 1.26 protein Pol; clone P10.21 protein Pol;
;   OTHER INFORMATION: clone P10.26 protein Pol
US-10-325-468-9

Alignment Scores:

Pred. No.:	1.38e-158	Length:	1003
Score:	3427.00	Matches:	638
Percent Similarity:	94.59%	Conservative:	44
Best Local Similarity:	88.49%	Mismatches:	33
Query Match:	74.94%	Indels:	6
DB:	16	Gaps:	3

US-09-610-313B-31 (1-2463) x US-10-325-468-9 (1-1003)

```
Qy      220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db        1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy      280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db       21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40

Qy      334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db       41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy      388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db       61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy      448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db       81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy      508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db      101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy      568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db      121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy      628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db      141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy      688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db      161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy      748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db      181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy      808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db      201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy      868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db      221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy      928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db      241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy      988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db      261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy     1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db      281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy     1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db      301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy     1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db      321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy     1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db      341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy     1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db      361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy     1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
              |||||||||||||||||   ||||||||||||:::|||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                 ||||||||||:::::::::||||||||:::||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              :::   ||||||||||||||||          ||||||||||||||||:::|||||||||
Db        481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              ||||||||||||||||||||||||||||||||||||   |||  ||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              ||||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
                 |||||||||:::|||||||||||||||||||||||||||||||||||||:::|||
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              |||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              |||||||||||||||||:::|||||||||||||||||||||   ||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::||||||||||||:::||||||||||||||||||||||||||||||||   |||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              ||||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::||||||||||||||||   |||||||||:::||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              :::
Db        721 Glu 721
```

RESULT 10
US-10-325-468-20

```
; Sequence 20, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
;  APPLICANT: Soong, Nay Wei
;  APPLICANT:  Pekrun, Katja
;  APPLICANT:  Shibata, Riri
;  TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
;  TITLE OF INVENTION:  MODELS OF HIV-1 PATHOGENESIS
;  FILE REFERENCE: 0166.210US
;  CURRENT APPLICATION NUMBER: US/10/325,468
;  CURRENT FILING DATE:  2002-12-19
;  PRIOR APPLICATION NUMBER: US 60/343,524
;  PRIOR FILING DATE: 2001-12-21
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone 1.27 protein Pol
US-10-325-468-20

Alignment Scores:
Pred. No.:              1.38e-158      Length:        1003
Score:                  3427.00        Matches:       638
Percent Similarity:     94.59%         Conservative:  44
Best Local Similarity:  88.49%         Mismatches:    33
Query Match:            74.94%         Indels:        6
DB:                     16             Gaps:          3

US-09-610-313B-31 (1-2463) x US-10-325-468-20 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                 ||||||||||||||      ||||||   ||||||      |||   :::|||
Db         21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCAGATCACC 387
                 ||||||||||||||:::||||||||||||:::      ||||||||||||||||
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              ||||||||||||||||||||:::|||||:::|||||||||:::|||||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              |||||||||||||||||||||||||||||:::|||   |||||||||:::||||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              :::|||||||||   ||||||:::||||||||||||:::|||||||||||||||||||||
Db        121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||||      ||||||   ||||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
```

```
Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              |||||||||||||||||||||||||||||||||||||||||||||::::::|||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              |||   |||||||||||||||      :::||||||:::|||:::|||||||||
Db        321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                    ||||||||||||||||||||||||||||||||||||      ||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::|||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1402 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1461
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1462 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1521
              ||||||||||||||||||   ||||||||||||:::||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1522 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1581
                 ||||||||||:::::::::||||||||:::||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy       1582 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1641
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1642 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1701
              :::   ||||||||||||||||||   |||||||||||||||||||||:::||||||||||
Db        481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy       1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1761
              |||||||||||||||||||||||||||||||||   |||   ||||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520

Qy       1762 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
              |||||||||||||||||||||||||||||||||      |||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1822 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
              |||||||||:::||||||||||||||||||||||||||||||||||||||||:::|||
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1882 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       2002 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2061
              |||||||||||||||:::||||||||||||||||||      ||||||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy       2062 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
              :::||||||||||||:::||||||||||||||||||||||||||||||   ||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2122 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2181
              ||||||||||||      ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2182 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2241
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2301
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2302 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2361
              ||||||:::||||||||||||||||||   ||||||:::||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy       2362 GAT 2364
              :::
Db        721 Glu 721
```

RESULT 11
US-10-296-734-2
; Sequence 2, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 995
;   TYPE: PRT
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: POL consensus polypeptide
US-10-296-734-2

```
Alignment Scores:
Pred. No.:              3.37e-158     Length:         995
Score:                  3419.00       Matches:        640
Percent Similarity:     94.44%        Conservative:   40
Best Local Similarity:  88.89%        Mismatches:     28
Query Match:            74.76%        Indels:         12
DB:                     15            Gaps:           4

US-09-610-313B-31 (1-2463) x US-10-296-734-2 (1-995)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::||||||||   ||||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
```

```
Db         21 ThrGlyAlaAsnSerSerAlaSerArgLysLeu---------GlyAspGly--------- 34

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACC---------CTGAACTTCCCCCAGATCACCCTG 390

Db         35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheProGlnIleThrLeu 53

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450

Db         54 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 73

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510

Db         74 ThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLys 93

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570

Db         94 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 113

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630

Db        114 IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690

Db        134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750

Db        154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 173

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810

Db        174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGluGlyLys 193

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870

Db        194 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 213

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930

Db        214 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990

Db        234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 253

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050

Db        254 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLys 273

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110

Db        274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 293

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170

Db        294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224

Db        314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTyrGlnTyrMetAspAsp 333

Qy       1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284

Db        334 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353

Qy       1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344

Db        354 AlaHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373

Qy       1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404

Db        374 PheLeuTrpMetGlyTyrGluLeuHisProAspArgTrpThrValGlnProIleGluLeu 393

Qy       1405 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1464

Db        394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413

Qy       1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1524

Db        414 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 433

Qy       1525 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1584

Db        434 LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuGluGluAsn 453

Qy       1585 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1644

Db        454 ArgGluIleLeuArgGluProValHisGlyValTyrTyrAspProSerLysAspLeuVal 473

Qy       1645 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1704

Db        474 AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPhe 493

Qy       1705 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1764

Db        494 LysAsnLeuLysThrGlyLysTyrSerArgLysArgSerAlaHisThrAsnAspValArg 513

Qy       1765 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1824

Db        514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533

Qy       1825 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884

Db        534 ProLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTyrTrp 553

Qy       1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1944

Db        554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 573

Qy       1945 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2004

Db        574 TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 593

Qy       2005 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2064

Db        594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613

Qy       2065 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2124

Db        614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633

Qy       2125 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2184

Db        634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653

Qy       2185 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2244

Db        654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673

Qy       2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2304

Db        674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693

Qy       2305 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364

Db        694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713
```

RESULT 12
US-10-296-734-1470
; Sequence 1470, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A

```
;  TITLE OF INVENTION: Synthetic molecules and uses therefor
;  FILE REFERENCE: Savine
;  CURRENT APPLICATION NUMBER: US/10/296,734
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: AU PQ7761/00
;  PRIOR FILING DATE: 2000-05-26
;  NUMBER OF SEQ ID NOS: 1507
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1470
;   LENGTH: 995
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus
US-10-296-734-1470

Alignment Scores:
Pred. No.:               1.61e-157      Length:        995
Score:                   3405.00        Matches:       638
Percent Similarity:      94.17%         Conservative:  40
Best Local Similarity:   88.61%         Mismatches:    30
Query Match:             74.46%         Indels:        12
DB:                      15             Gaps:          4

US-09-610-313B-31 (1-2463) x US-10-296-734-1470 (1-995)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
Db         21 ThrGlyAlaAsnSerSerAlaSerArgLysLeu---------GlyAspGly--------- 34

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACC---------CTGAACTTCCCCCAGATCACCCTG 390
Db         35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheProGlnIleThrLeu 53

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
Db         54 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 73

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
Db         74 ThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLys 93

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
Db         94 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 113

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
Db        114 IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
Db        134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
Db        154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 173

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
Db        174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGluGlyLys 193

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
Db        194 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 213

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
Db        214 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
Db        234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 253

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
Db        254 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLys 273

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
Db        274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 293

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
Db        294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
Db        314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTyrGlnTyrMetAspAsp 333

Qy       1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
Db        334 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353

Qy       1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
Db        354 AlaHisLeuLeuSerTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373

Qy       1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404
Db        374 PheLeuTrpMetGlyTyrGluLeuHisProAspArgTrpThrValGlnProIleGluLeu 393

Qy       1405 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1464
Db        394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413

Qy       1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1524
Db        414 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 433

Qy       1525 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1584
Db        434 LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuGluGluAsn 453

Qy       1585 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1644
Db        454 ArgGluIleLeuArgIleProValHisGlyValTyrTyrAspProSerLysAspLeuVal 473

Qy       1645 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1704
Db        474 AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPhe 493

Qy       1705 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1764
Db        494 LysAsnLeuLysThrGlyLysTyrSerArgLysArgSerAlaHisThrAsnAspValArg 513

Qy       1765 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1824
Db        514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533

Qy       1825 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884
Db        534 ProLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTyrTrp 553

Qy       1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1944
Db        554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 573

Qy       1945 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 2004
Db        574 TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 593

Qy       2005 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2064
```

```
Db          594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613

Qy         2065 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2124
                :::|||||||||||||||||||||||||||||||||       ||||||    |||||||||
Db          614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633

Qy         2125 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2184
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653

Qy         2185 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2244
                ||||||||||||||:::|||||||||||:::|||:::|||||||||||||:::||||||
Db          654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673

Qy         2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2304
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693

Qy         2305 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2364
                |||:::||||||||||||      |||||||||||||||||||||||||||||||||:::
Db          694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713
```

RESULT 13
US-10-296-734-1471
; Sequence 1471, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1471
;   LENGTH: 1006
;   TYPE: PRT
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: POL Consensus A
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (15)..(15)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (22)..(23)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (26)..(26)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (30)..(30)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (36)..(36)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (38)..(39)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (41)..(44)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (46)..(46)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (48)..(48)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (51)..(51)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (53)..(55)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (58)..(58)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (73)..(73)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (75)..(75)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (105)..(105)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (130)..(130)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (173)..(173)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (194)..(194)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (197)..(197)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (255)..(255)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (280)..(281)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (293)..(293)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (300)..(300)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (316)..(316)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (320)..(320)
;   OTHER INFORMATION: X = unknown amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (331)..(333)

```
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (335)..(335)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (365)..(365)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (369)..(369)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (373)..(373)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (402)..(403)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (434)..(434)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (439)..(439)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (452)..(452)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (470)..(470)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (480)..(480)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (487)..(487)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (514)..(514)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (534)..(535)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (544)..(544)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (553)..(553)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (558)..(558)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (593)..(593)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (610)..(610)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (675)..(675)
;    OTHER INFORMATION: X = unknown amino acid
;    FEATURE:

Alignment Scores:
Pred. No.:              4.62e-150    Length:        1006
Score:                  3251.50      Matches:       618
Percent Similarity:     88.81%       Conservative:  25
Best Local Similarity:  85.36%       Mismatches:    72
Query Match:            71.10%       Indels:        9
DB:                     15           Gaps:          3

US-09-610-313B-31 (1-2463) x US-10-296-734-1471 (1-1006)

Qy      220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db        1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArg***PheSerSerGluGln 20

Qy      280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCC------ 333
Db       21 Thr******AsnSer***ThrSerArg***LeuTrpAspGlyGly***Asp******Leu 40

Qy      334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGC---------------ACCCTGAACTTCCCC 378
Db       41 ************Gly***Glu***GlnGly***Gly*********SerPhe***PhePro 60

Qy      379 CAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAG 438
Db       61 GlnIleThrLeuTrpGlnArgProLeuValThrVal***Ile***GlyGlnLeuIleGlu 80

Qy      439 GCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAG 498
Db       81 AlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLys 100

Qy      499 TGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAG 558
Db      101 TrpLysProLys***IleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGln 120

Qy      559 ATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCC 618
Db      121 IleLeuIleGluIleCysGlyLysLys***IleGlyThrValLeuValGlyProThrPro 140

Qy      619 GTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATC 678
Db      141 ValAsnIleIleGlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIle 160

Qy      679 AGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAG 738
Db      161 SerProIleGluThrValProValLysLeuLysPro***MetAspGlyProLysValLys 180

Qy      739 CAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAG 798
Db      181 GlnTrpProLeuThrGluGluLysIleLysAlaLeuThr***IleCys***GluMetGlu 200

Qy      799 AAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCC 858
Db      201 LysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAla 220

Qy      859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918
Db      221 IleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys 240

Qy      919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAG 978
Db      241 ArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis***AlaGlyLeuLysLys 260

Qy      979 AAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAG 1038
Db      261 LysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAsp*** 280

Qy     1039 GACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATC 1098
```

```
Db         281 ***PheArgLysTyrThrAlaPheThrIleProSer***AsnAsnGluThrProGly*** 300

Qy        1099 CGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGC 1158
Db         301 ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro***IlePheGln*** 320

Qy        1159 AGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG 1218
Db         321 SerMetThrLysIleLeuGluProPheArg*********Pro***IleValIleTyrGln 340

Qy        1219 ------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATC 1272
Db         341 TyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIle 360

Qy        1273 GAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAG 1332
Db         361 GluGluLeuArg***HisLeuLeu***TrpGlyPhe***ThrProAspLysLysHisGln 380

Qy        1333 AAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAG 1392
Db         381 LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGln 400

Qy        1393 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1452
Db         401 Pro******LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 420

Qy        1453 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1512
Db         421 LysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys***LysGlnLeuCys***Leu 440

Qy        1513 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1572
Db         441 LeuArgGlyAlaLysAlaLeuThrAspIleVal***LeuThrGluGluAlaGluLeuGlu 460

Qy        1573 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1632
Db         461 LeuAlaGluAsnArgGluIleLeuLys***ProValHisGlyValTyrTyrAspPro*** 480

Qy        1633 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1692
Db         481 LysAspLeuValAlaGlu***GlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyr 500

Qy        1693 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1752
Db         501 GlnGluProPheLysAsnLeuLysThrGlyLysTyrAla***LysArgSerAlaHisThr 520

Qy        1753 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1812
Db         521 AsnAspValLysGlnLeuThrGluValValGlnLysVal******GluSerIleValIle 540

Qy        1813 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1872
Db         541 TrpGlyLys***ProLysPheArgLeuProIleGln***GluThrTrpGlu***TrpTrp 560

Qy        1873 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1932
Db         561 MetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu 580

Qy        1933 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1992
Db         581 ValLysLeuTrpTyrGlnLeuGluLysAspProIle***GlyAlaGluThrPheTyrVal 600

Qy        1993 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2052
Db         601 AspGlyAlaAlaAsnArgGluThrLys***GlyLysAlaGlyTyrValThrAspArgGly 620

Qy        2053 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2112
Db         621 ArgGlnLysValValSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIle 640

Qy        2113 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2172
Db         641 HisLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAla 660

Qy        2173 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2232
Db         661 LeuGlyIleIleGlnAlaGlnProAspArgSerGluSerGlu***ValAsnGlnIleIle 680

Qy        2233 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2292
Db         681 GluLysLeuIle***Lys***LysValTyrLeuSerTrpValProAlaHisLysGlyIle 700

Qy        2293 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2352
Db         701 GlyGlyAsnGluGlnValAspLysLeuValSer***GlyIleArgLysValLeuPheLeu 720

Qy        2353 GACGGCATCGAT 2364
Db         721 AspGlyIleAsp 724
```

RESULT 14
US-10-301-661A-6
; Sequence 6, Application US/10301661A
; Publication No. US20030157660A1
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-
; APPLICANT: INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAUCLERE, Philippe
; APPLICANT: LOUSSERT-AJAKA, Ibtissam
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSSI, Francoise
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/10/301,661A
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US/09/319,588C
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1014
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-10-301-661A-6

Alignment Scores:

Pred. No.:	9.3e-147	Length:	1014
Score:	3183.50	Matches:	583
Percent Similarity:	90.25%	Conservative:	74
Best Local Similarity:	80.08%	Mismatches:	58
Query Match:	69.62%	Indels:	13
DB:	14	Gaps:	4

US-09-610-313B-31 (1-2463) x US-10-301-661A-6 (1-1014)

```
Qy         220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db           1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProProAspAsn 20

Qy         280 AAC------CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG--------------- 318
Db          21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40

Qy         319 ---CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG---------GGCACC 366
Db          41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60

Qy         367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db          61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80
```

```
Qy     427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db      81 GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluGluLeuGln 100

Qy     487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db     101 LeuGluGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArg 120

Qy     547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db     121 GlnTyrAspAsnIleThrValAspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140

Qy     607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db     141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160

Qy     667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db     161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180

Qy     727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db     181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200

Qy     787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db     201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220

Qy     847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db     221 ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240

Qy     907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db     241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260

Qy     967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db     261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280

Qy    1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db     281 ProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGlu 300

Qy    1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db     301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320

Qy    1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db     321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340

Qy    1207 GTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db     341 IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360

Qy    1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db     361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380

Qy    1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db     381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400

Qy    1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db     401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420

Qy    1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db     421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440

Qy    1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db     441 LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460

Qy    1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db     461 AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480

Qy    1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db     481 TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThr 500

Qy    1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db     501 TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520

Qy    1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db     521 SerAlaHisThrAsnAspIleLysGlnLeuValGluValValArgLysValAlaThrGlu 540

Qy    1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db     541 SerIleValIleTrpGlyLysThrProLysPheArgLeuProValGlnLysGluValTrp 560

Qy    1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db     561 GluAlaTrpTrpThrAspHisTrpGlnAlaThrTrpIleProGluTrpGluPheValAsn 580

Qy    1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db     581 ThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGluProIleSerGlyAlaGlu 600

Qy    1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db     601 ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620

Qy    2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db     621 ThrAspArgGlyArgGlnLysValValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640

Qy    2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db     641 LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660

Qy    2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db     661 SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680

Qy    2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db     681 SerGlnIleIleGluGluLeuIleLysLysGluArgValTyrLeuSerTrpValProAla 700

Qy    2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db     701 HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720

Qy    2341 GTGCTGTTCCTGGACGGCATCGAT 2364
Db     721 IleLeuPheLeuAspGlyIleGlu 728
```

RESULT 15
US-09-952-060-35
; Sequence 35, Application US/09952060
; Publication No. US20030044421A1
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.
; APPLICANT: Youil, Rima
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.

```
;  TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
;  TITLE OF INVENTION:  VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
;  TITLE OF INVENTION:  MODIFICATIONS
;  FILE REFERENCE: 20747Y
;  CURRENT APPLICATION NUMBER: US/09/952,060
;  CURRENT FILING DATE:  2001-09-14
;  PRIOR APPLICATION NUMBER: PCT/US01/28861
;  PRIOR FILING DATE: 2001-09-14
;  PRIOR APPLICATION NUMBER: 60/317,814
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: 60/279,056
;  PRIOR FILING DATE: 2001-03-27
;  PRIOR APPLICATION NUMBER: 60/233,180
;  PRIOR FILING DATE: 2000-09-15
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 1350
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Codon optimized gag-IA pol fusion
US-09-952-060-35

Alignment Scores:
Pred. No.:              1.22e-143     Length:         1350
Score:                  3119.50       Matches:        599
Percent Similarity:     81.16%        Conservative:   43
Best Local Similarity:  75.73%        Mismatches:     59
Query Match:            68.22%        Indels:         90
DB:                     10            Gaps:           6

US-09-610-313B-31 (1-2463) x US-09-952-060-35 (1-1350)

Qy     13 ATGGCCGAGGCCATGAGCCAGGCCACC---AGCGCCAACATCCTGATGCAGCGCAGCAAC 69
Db    363 LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn 382

Qy     70 TTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCC 129
Db    383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysValGlyHisIleAla 402

Qy    130 CGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAG 189
Db    403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422

Qy    190 ATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAG 249
Db    423 MetLysAspCysAsnGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProSerHisLy 442

Qy    250 GGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAG 309
Db    442 sGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProGluGluSe 462

Qy    310 CTGCAGGTGCGCG------GCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGC 363
Db    462 rPheArgPheGlyGluGluLysThrThrProSerGlnLysGlnGluProIleAspLys-- 481

Qy    364 ACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGC 423
Db    482 GluLeu--------------------------TyrProLeuAlaSerLeuArg------ 490

Qy    424 GGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATG 483
Db    490 ------------------------------------------------------------ 490

Qy    484 AGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTG 543
Db    490 ------------------------------------------------------------ 490

Qy    544 CGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTG 603
Db    490 ------------------------------------------------------------ 490

Qy    604 ATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACC 663
Db    491 -----------------SerLeuPheGlyAsnAspProSerSerGlnMetAla------ 502

Qy    664 CTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGAC 723
Db    503 ---------ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAsp 519

Qy    724 GGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATC 783
Db    520 GlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIle 539

Qy    784 TGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAAC 843
Db    540 CysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsn 559

Qy    844 ACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTC 903
Db    560 ThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPhe 579

Qy    904 CGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCC 963
Db    580 ArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPro 599

Qy    964 GCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGC 1023
Db    600 AlaGlyLeuLysLysLysLysSerValThrValLeuAlaValGlyAspAlaTyrPheSer 619

Qy   1024 GTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAAC 1083
Db    620 ValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsn 639

Qy   1084 GAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCC 1143
Db    640 GluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro 659

Qy   1144 AGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAG 1203
Db    660 AlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAsp 679

Qy   1204 ATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAG 1257
Db    680 IleValIleTyrGlnTyrMetAlaAlaLeuTyrValGlySerAspLeuGluIleGlyGln 699

Qy   1258 CACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCC 1317
Db    700 HisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrPro 719

Qy   1318 GACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGAC 1377
Db    720 AspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAsp 739

Qy   1378 AAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATC 1437
Db    740 LysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIle 759

Qy   1438 CAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGC 1497
Db    760 GlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArg 779

Qy   1498 CAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAG 1557
Db    780 GlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGlu 799

Qy   1558 GAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTG 1617
Db    800 GluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyVal 819

Qy   1618 TACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGG 1677
Db    820 TyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrp 839
```

```
Qy       1678 ACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATG 1737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        840 ThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859

Qy       1738 CGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATG 1797
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||      
Db        860 ArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr 879

Qy       1798 GAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACC 1857
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        880 GluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGluThr 899

Qy       1858 TGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTG 1917
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        900 TrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVal 919

Qy       1918 AACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCC 1977
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        920 AsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAla 939

Qy       1978 GAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTAC 2037
              |||||||||||||||   |||||||||||||||||||||||||:::||||||||||||||
Db        940 GluThrPheTyrValAlaGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyr 959

Qy       2038 GTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACC 2097
              ||||||:::|||||||||||||||:::|||:::||||||:::||||||||||||||||||
Db        960 ValThrAsnArgGlyArgGlnLysValValThrLeuThrAspThrThrAsnGlnLysThr 979

Qy       2098 GAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACC 2157
                 |||||||||||||||   ||||||||||||||||||   ||||||||||||||||||
Db        980 AlaLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThr 999

Qy       2158 GACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTG 2217
                 |||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db       1000 AlaSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeu 1019

Qy       2218 GTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCC 2277
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       1020 ValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValPro 1039

Qy       2278 GCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGC 2337
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||   |||||||||
Db       1040 AlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArg 1059

Qy       2338 AAGGTGCTGTTCCTGGACGGCATCGAT 2364
              |||||||||||||||||||||||||||
Db       1060 LysValLeuPheLeuAspGlyIleAsp 1068
```

Search completed: June 2, 2005, 05:28:17
Job time : 228.167 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  1, 2005, 11:33:41 ; Search time 167.424 Seconds
                                          (without alignments)
                                          11351.628 Million cell updates/sec

Title:          US-09-610-313B-32
Perfect score:  4555
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2457

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters:      4211384

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155136_15139/app_query.fasta_1.7893
-DB=A_Geneseq_16Dec04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_557_@runat_31052005_155136_15139 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      A_Geneseq_16Dec04:*
                1:  geneseqp1980s:*
                2:  geneseqp1990s:*
                3:  geneseqp2000s:*
                4:  geneseqp2001s:*
                5:  geneseqp2002s:*
                6:  geneseqp2003as:*
                7:  geneseqp2003bs:*
                8:  geneseqp2004s:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                        Description
 ----------------------------------------------------------------------------
    1  3723.5   81.7   1435  8  ADS19486                 Ads19486 Modified
    2  3701.5   81.3   3183  8  ADP84803                 Adp84803 HIV-1 hyb
    3  3668.5   80.5   1457  8  ADN36406                 Adn36406 Human pro
    4    3577   78.5   1003  3  AAB69289                 Aab69289 HIV-1 non
    5    3568   78.3    999  3  AAB69286                 Aab69286 HIV-1 non
    6    3563   78.2   1005  3  AAB69287                 Aab69287 HIV-1 non
    7    3527   77.4    998  6  AAE37601                 Aae37601 HIV-1 sub
    8    3520   77.3   3025  4  AAB86169                 Aab86169 HIV-1 sub
    9    3478   76.4    854  5  AAM48949                 Aam48949 HIV-1 sub
   10  3474.5   76.3   1000  3  AAB69282                 Aab69282 HIV-1 non
   11  3456.5   75.9   1002  2  AAW72993                 Aaw72993 HIV isola
   12    3454   75.8   1003  6  AAO30963                 Aao30963 HIV pol p
   13    3454   75.8   1003  8  ADN36413                 Adn36413 HIV prote
   14  3451.5   75.8   1002  1  AAP81861                 Aap81861 Sequence
   15    3451   75.8   1003  6  ABR55489                 Abr55489 Amino aci
   16    3449   75.7   1003  1  AAP60420                 Aap60420 Sequence
   17    3449   75.7   1003  1  AAP70861                 Aap70861 Sequence
   18    3443   75.6   1003  3  AAY70602                 Aay70602 Codon opt
   19    3443   75.6   1003  3  AAY70601                 Aay70601 Corrected
   20    3441   75.5   1003  5  AAO19387                 Aao19387 Lymphaden
   21  3435.5   75.4   1003  2  AAR08060                 Aar08060 HIV-1 pol
   22    3432   75.3   1015  1  AAP60347                 Aap60347 HTLV-III
   23    3432   75.3   1015  2  AAR43867                 Aar43867 HTLV-III
   24    3432   75.3   1015  4  AAB85993                 Aab85993 Amino aci
   25    3429   75.3   1003  3  AAB10047                 Aab10047 HIV-1 pol
   26    3429   75.3   1003  3  AAY70600                 Aay70600 Wild type
   27    3428   75.3   1016  2  AAR08062                 Aar08062 AcNPV-HIV
   28    3425   75.2   1015  2  AAR43875                 Aar43875 HTLV-III
   29    3421   75.1   1016  2  AAR08063                 Aar08063 HIV-1 pol
   30    3418   75.0   2033  2  AAR08056                 Aar08056 HIV-1 pol
   31    3414   75.0   1016  2  AAR08054                 Aar08054 HIV-1 pol
   32    3412   74.9   1003  3  AAB69284                 Aab69284 HIV-1 non
   33    3411   74.9   2033  2  AAR08055                 Aar08055 HIV-1 pol
   34    3410   74.9   1003  8  ADP20076                 Adp20076 Human imm
   35    3406   74.8    739  5  AAU11874                 Aau11874 HIV pol p
   36    3406   74.8   1003  1  AAP61508                 Aap61508 Sequence
   37    3406   74.8   1003  2  AAR29705                 Aar29705 pol gene
   38    3406   74.8   1003  3  AAY77301                 Aay77301 HIV-1 (AT
   39    3406   74.8   1003  5  AAE35789                 Aae35789 ARV-2 (9B
   40    3406   74.8   1003  6  ABU63185                 Abu63185 Pol prote
   41    3406   74.8   1491  1  AAP91048                 Aap91048 Transcrip
   42    3403   74.7   1003  3  AAB69279                 Aab69279 HIV-1 non
   43    3401   74.7   1003  6  ABU57552                 Abu57552 AIDS asso
   44    3399   74.6   1012  6  ABU63325                 Abu63325 Human lym
   45  3396.5   74.6   1003  2  AAR08059                 Aar08059 HIV-1 pol

                              ALIGNMENTS

RESULT 1
ADS19486
ID   ADS19486 standard; protein; 1435 AA.
XX
AC   ADS19486;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Modified HIV-1 GagPol protein.
XX
KW   Gene delivery; epithelial cell; respiratory system; mammal;
KW   lentiviral packaging system; expression vector; gagpol; gene expression;
KW   lentivirus; cystic fibrosis transmembrane conductance regulator; CFTR;
KW   cystic fibrosis; CF; deltaPD; transepithelial potential difference;
KW   gene therapy; HIV-1; mutant; mutein.
XX
OS   Human immunodeficiency virus 1.
OS   Synthetic.
XX
PN   US2004037780-A1.
XX
PD   26-FEB-2004.
XX
PF   23-AUG-2002; 2002US-00226638.
XX
PR   19-NOV-2001; 2001AU-00008942.
XX
PA   (PARS/) PARSONS D.
PA   (ANSO/) ANSON D.
PA   (LIMB/) LIMBERIS M.
PA   (FULL/) FULLER M.
XX
PI   Parsons D,  Anson D,  Limberis M,  Fuller M;
XX
```

```
DR   WPI; 2004-191004/18.
DR   N-PSDB; ADS19485.
XX
PT   Delivering exogenous (e.g. CFTR) genes to epithelial cells in the
PT   respiratory tract using a penetration agent and a lentiviral packaging
PT   system, useful for treating e.g. cystic fibrosis.
XX
PS   Disclosure; Fig 16; 45pp; English.
XX
CC   The present invention relates to a method of delivering one or more
CC   exogenous genes for expression in an epithelial cell in the respiratory
CC   system of a mammal to give persistent expression of the gene in the
CC   epithelial cell, using a lentiviral packaging (expression vector) system.
CC   Also disclosed are (i) a recombinant lentiviral packaging system ,
CC   comprising a first nucleic acid molecule including a gag gene sequence
CC   encoding a Gag protein, and a second nucleic acid molecule including a
CC   gagpol gene sequence encoding a GagPol protein (the gagpol gene sequence
CC   has degenerative nucleotide changes in the frame shift sequence AUUUUUU
CC   to reduce the chance of a frame shift which switches expression of the
CC   Gagpol protein to the Gag protein in wild type lentivirus and the
CC   packaging system additionally comprises a lentiviral vector nucleic acid
CC   molecule which does not encode either the gag gene or the gagpol gene or
CC   both, and (ii) a recombinant nucleic acid molecule encoding a lentiviral
CC   gagpol gene having degenerative nucleotide changes in the frame shift
CC   sequence AUUUUUU to reduce the chance of a frame shift which switches
CC   expression of the Gagpol protein to the Gag protein in the wild type
CC   lentivirus. The exogenous gene is expressed in sufficient numbers of
CC   cells and amounts to provide an ameliorating effect for a respiratory
CC   condition. The exogenous gene is cystic fibrosis transmembrane
CC   conductance regulator (CFTR) and the condition is cystic fibrosis
CC   ((CF)the expression of the CFTR gene is sufficient to provide a
CC   significant shift of a reduced deltaPD (transepithelial potential
CC   difference) back to normal levels in the mammal). The cell is non-
CC   terminally differentiated and is capable of differentiating into 2 or
CC   more cell classes e.g. ciliated cells, non-ciliated cells, secretory
CC   cells and basal cells. The exogenous gene may be enzymic. The gagpol gene
CC   sequence has additionally degenerate nucleotide substitutions which
CC   destabilise the hairpin structure associated with the frame shift event.
CC   The gag gene is a truncation of the wild type gagpol gene so that it can
CC   no longer be translated to form Gagpol. The lentivirus is HIV or HIV
CC   derived. The method and recombinant lentiviral packaging system may be
CC   used for delivering nucleic acids to epithelial cells in the respiratory
CC   system of a mammal. In particular they may be used to deliver the CFTR
CC   gene for the treatment of CF. It has been found that the delivery of a
CC   recombinant lentivirus carrying an exogenous gene to the respiratory
CC   system, following the delivery of a non-toxic amount of a penetration
CC   enhancer can provide persistence of expression of a gene product. The
CC   present sequence represents modified HIV-1 GagPol protein.
XX
SQ   Sequence 1435 AA;

Alignment Scores:
Pred. No.:              6.59e-207     Length:        1435
Score:                  3723.50       Matches:       698
Percent Similarity:     93.81%        Conservative:  44
Best Local Similarity:  88.24%        Mismatches:    40
Query Match:            81.75%        Indels:        9
DB:                     8             Gaps:          5

US-09-610-313B-32 (1-2457) x ADS19486 (1-1435)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACC---AGCGCCAACATCCTGATGCAGCGCAGCAAC 69
              :::||||||||||||||||   |||   ||||||  |||:::||||||||    |||
Db        363 LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn 382

Qy         70 TTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCC 129
              |||:::      ::::::   :::|||||||||||||||||||||||||||||||||
Db        383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla 402

Qy        130 CGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAG 189
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422

Qy        190 ATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAG 249
              |||||||||||||||||||||||||||||||    ||||||||||||||||||||||
Db        423 MetLysAspCysThrGluArgGlnAlaAsnPheLeuArgGluAspLeuAlaPheProGln 442

Qy        250 GGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAG 309
              |||||||||||:::|||  ||||||||  |||||||||||||||        ||||||
Db        443 GlyLysAlaArgLysPheSerSerGluGlnThrArgAlaAsnSerProIleArgArgGlu 462

Qy        310 CTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGC 363
              ||||||       |||   :::|||     |||||||||||||:::||||||||
Db        463 ArgGlnValTrpArgArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGly 482

Qy        364 ACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAG 417
              |||:::      :::||||||||||||||||||||||||||||||||||:::||||||
Db        483 ThrValSerPheSerPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLys 502

Qy        418 GTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAG 477
              :::|||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        503 IleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGlu 522

Qy        478 GAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATC 537
              ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        523 GluMetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIle 542

Qy        538 AAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACC 597
              ||||||||||||||||||||||||       |||||||||||||      ||||||||
Db        543 LysValArgGlnTyrAspGlnIleProIleGluIleCysGlyHisLysAlaIleGlyThr 562

Qy        598 GTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGC 657
              ||||||:::|||||||||||||||||||||||||||||||:::|||||||||:::|||
Db        563 ValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGly 582

Qy        658 TGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGC 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 CysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGly 602

Qy        718 ATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACC 777
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 MetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuVal 622

Qy        778 GCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCC 837
                 ||||||   ||||||||||||||||||||:::||||||||||||||||||||||||
Db        623 GluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnPro 642

Qy        838 TACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTG 897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 TyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuVal 662

Qy        898 GACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC 957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 AspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIlePro 682

Qy        958 CACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTAC 1017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 HisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyr 702

Qy       1018 TTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATC 1077
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        703 PheSerValProLeuHisGluAspPheArgLysTyrThrAlaPheThrIleProSerIle 722

Qy       1078 AACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGC 1137
              ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db        723 AsnAsnGluThrProGlyThrArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGly 742

Qy       1138 AGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAAC 1197
              ||||||:::|||||||||||||||||||||   |||||||||||||||||||    :::|||
Db        743 SerProAlaIlePheGlnSerSerMetThrThrIleLeuGluProPheArgLysGlnAsn 762

Qy       1198 CCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATC 1251
              |||::::::||||||||||||            |||||||||||||||||||||||||||
Db        763 ProAspLeuValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIle 782

Qy       1252 GGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACC 1311
```

```
Db        783 GlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThr 802

Qy       1312 ACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCAC 1365
Db        803 ThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHis 822

Qy       1366 CCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAAC 1425
Db        823 ProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsn 842

Qy       1426 GACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAG 1485
Db        843 AspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys 862

Qy       1486 GTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTG 1545
Db        863 ValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeu 882

Qy       1546 ACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCAC 1605
Db        883 ThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHis 902

Qy       1606 GGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGAC 1665
Db        903 GlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGly 922

Qy       1666 CAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCC 1725
Db        923 GlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAla 942

Qy       1726 AAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATC 1785
Db        943 ArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIle 962

Qy       1786 GCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAG 1845
Db        963 AlaThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLys 982

Qy       1846 GAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAG 1905
Db        983 GluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGlu 1002

Qy       1906 TTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATC 1965
Db       1003 PheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIle 1022

Qy       1966 GGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCC 2025
Db       1023 GlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAla 1042

Qy       2026 GGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAG 2085
Db       1043 GlyTyrValThrAsnLysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGln 1062

Qy       2086 AAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATC 2145
Db       1063 LysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIle 1082

Qy       2146 GTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGC 2205
Db       1083 ValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspArgSerGluSer 1102

Qy       2206 GAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGG 2265
Db       1103 GluLeuValSerGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrp 1122

Qy       2266 GTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGC 2325
Db       1123 ValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGly 1142

Qy       2326 ATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db       1143 IleArgLysValLeuPheLeuAspGlyIleAsp 1153
```

```
RESULT 2
ADP84803
ID   ADP84803 standard; protein; 3183 AA.
XX
AC   ADP84803;
XX
DT   23-SEP-2004  (first entry)
XX
DE   HIV-1 hybrid isolate pNL4-3 protein.
XX
KW   eliciting; inducing; immune response; HIV; antigen; non-pathogenic;
KW   vaccination; vaccine; HIV-I.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO2004056391-A1.
XX
PD   08-JUL-2004.
XX
PF   19-DEC-2003; 2003WO-AU001705.
XX
PR   20-DEC-2002; 2002AU-00953556.
PR   17-SEP-2003; 2003AU-00905067.
XX
PA   (UYNE-) UNIV NEW SOUTH WALES.
XX
PI   Kent SJ,  Purcell DF,  Boyle DB,  Ramsay A,  Thomson S,  Ramshaw IA;
XX
DR   WPI; 2004-500267/47.
DR   N-PSDB; ADP84804.
XX
PT   Eliciting or inducing in a mammal an immune response against HIV-I
PT   subtype AE, B or C by administering sequential doses of a recombinant
PT   plasmid and viral vectors containing the nucleic acid molecules encoding
PT   the HIV antigens.
XX
PS   Disclosure; Fig 12; 280pp; English.
XX
CC   The invention relates to a novel method for eliciting or inducing in a
CC   mammal an immune response directed to a virus, preferably Human
CC   immunodeficiency virus (HIV). The method comprises sequentially
CC   administering to the mammal one or more sequential doses of a recombinant
CC   plasmid vector or recombinant viral vector or its derivative, into which
CC   the nucleic acid molecules encoding all, part or a modified form of two
CC   or more antigens of the virus are incorporated, and one or more optimized
CC   CpG motifs, where the antigens have been rendered substantially non-
CC   pathogenic. The invention further comprises: a method for treating or
CC   preventing viral infection in a mammal; a method of vaccinating a mammal
CC   against a viral pathogen; a method of eliciting or inducing, in a mammal,
CC   an immune response directed to HIV; a vaccine capable of inducing an
CC   immune response directed to a virus comprising the recombinant plasmid
CC   vector or recombinant viral vector or its functional derivative; a
CC   nucleic acid construct or its functional derivative comprising the
CC   plasmid vector or recombinant viral vector; a plasmid vector; a
CC   recombinant viral vector; a pharmaceutical composition comprising the
CC   nucleic acid, plasmid vector or recombinant viral vector constructs. The
CC   method is useful in eliciting or inducing in a mammal an immune response
CC   against HIV-I subtype AE, B or C. This sequence represents the protein of
CC   the HIV-1 hybrid isolate pNL4-3 DNA of the invention.
XX
SQ   Sequence 3183 AA;

Alignment Scores:
Pred. No.:              1.47e-205     Length:        3183
Score:                  3701.50       Matches:       703
Percent Similarity:     86.96%        Conservative:  44
Best Local Similarity:  81.84%        Mismatches:    35
Query Match:            81.26%        Indels:        77
DB:                     8             Gaps:          7

US-09-610-313B-32 (1-2457) x ADP84803 (1-3183)
```

```
Qy     13 ATGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCCTGATGCAGCGCAGCAAC 69
Db    363 LeuAlaGluAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLysGlyAsn 382

Qy     70 TTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCC 129
Db    383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla 402

Qy    130 CGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAG 189
Db    403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422

Qy    190 ATGAAGGACTGCACCGAGCGCCAGGCCAAC------------------------------ 219
Db    423 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys 442

Qy    220 ---------------------------------------------TTCTTCCGCGAGGAC 234
Db    443 GlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPhePheArgGluAsp 462

Qy    235 CTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGC 294
Db    463 LeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSer 482

Qy    294 ------------------------------------------------------------ 294
Db    483 ProProGluGluSerPheArgPheGlyGluGluThrThrThrProSerGlnLysGlnGlu 502

Qy    295 ------------------------------CCCACCAGCCGCGAGCTGCAGGTG------ 318
Db    503 ProIleAspLysGluLeuTyrProLeuAlaProThrArgArgGluLeuGlnValTrpGly 522

Qy    319 CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC------ 372
Db    523 ArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSer 542

Qy    373 ---------------------------------------TTCCCCCAGATCACCCTGTGG 393
Db    543 SerLeuArgSerLeuPheGlySerAspProSerSerGlnPheProGlnIleThrLeuTrp 562

Qy    394 CAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACC 453
Db    563 GlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThr 582

Qy    454 GGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATG 513
Db    583 GlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysProLysMet 602

Qy    514 ATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATC 573
Db    603 IleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIle 622

Qy    574 TGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGC 633
Db    623 CysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGly 642

Qy    634 CGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACC 693
Db    643 ArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThr 662

Qy    694 GTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACC 753
Db    663 ValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThr 682

Qy    754 GAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATC 813
Db    683 GluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIle 702

Qy    814 ACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGAC 873
Db    703 SerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAsp 722

Qy    874 AGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTC 933
Db    723 SerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPhe 742

Qy    934 TGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACC 993
Db    743 TrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSerValThr 762

Qy    994 GTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTAC 1053
Db    763 ValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyr 782

Qy   1054 ACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAAC 1113
Db    783 ThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsn 802

Qy   1114 GTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATC 1173
Db    803 ValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThrLysIle 822

Qy   1174 CTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTG 1227
Db    823 LeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeu 842

Qy   1228 TACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAG 1287
Db    843 TyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGln 862

Qy   1288 CACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTC 1347
Db    863 HisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPhe 882

Qy   1348 CTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCC 1401
Db    883 LeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleValLeuPro 902

Qy   1402 GAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCC 1461
Db    903 GluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAla 922

Qy   1462 AGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAG 1521
Db    923 SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLys 942

Qy   1522 GCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGC 1581
Db    943 AlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArg 962

Qy   1582 GAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCC 1641
Db    963 GluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAla 982

Qy   1642 GAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAG 1701
Db    983 GluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLys 1002

Qy   1702 AACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAG 1761
Db   1003 AsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspValLysGln 1022

Qy   1762 CTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCC 1821
Db   1023 LeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThrPro 1042

Qy   1822 AAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAG 1881
Db   1043 LysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrpGln 1062

Qy   1882 GCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTAC 1941
Db   1063 AlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyr 1082

Qy   1942 CAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAAC 2001
```

```
Db       1083 GlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAlaAsn 1102

Qy       2002 CGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTG 2061
Db       1103 ArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValVal 1122

Qy       2062 AGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAG 2121
Db       1123 ProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGln 1142

Qy       2122 GACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAG 2181
Db       1143 AspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGln 1162

Qy       2182 GCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAG 2241
Db       1163 AlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleLys 1182

Qy       2242 AAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAG 2301
Db       1183 LysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGln 1202

Qy       2302 ATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db       1203 ValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1221
```

RESULT 3
ADN36406

```
ID   ADN36406 standard; protein; 1457 AA.
XX
AC   ADN36406;
XX
DT   15-JUL-2004  (first entry)
XX
DE   Human protein for anti-HIV vaccine.
XX
KW   anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX
OS   Homo sapiens.
XX
PN   WO2004035006-A2.
XX
PD   29-APR-2004.
XX
PF   17-OCT-2003; 2003WO-US033112.
XX
PR   18-OCT-2002; 2002US-0419465P.
XX
PA   (AARO-) AARON DIAMOND AIDS RES CENT.
XX
PI   Huang Y,  Ho DD,  Chen Z;
XX
DR   WPI; 2004-348328/32.
DR   N-PSDB; ADN36405.
XX
PT   Nucleic acid vector comprising at least one HIV sequence operably linked
PT   to a promoter and encoding a protein that does not assemble into viral
PT   particles, useful in immunizing a subject against HIV infection.
XX
PS   Disclosure; SEQ ID NO 20; 166pp; English.
XX
CC   The invention relates to a nucleic acid vector comprising at least one
CC   HIV sequence operably linked to a promoter and encoding a protein that
CC   does not assemble into viral particles. The nucleic acid vector is useful
CC   in immunizing a subject against HIV infection. This sequence corresponds
CC   to a peptide used in the invention.
XX
SQ   Sequence 1457 AA;

Alignment Scores:
Pred. No.:              1e-203        Length:         1457
Score:                  3668.50       Matches:        702
Percent Similarity:     91.36%        Conservative:   28
Best Local Similarity:  87.86%        Mismatches:     47
Query Match:            80.54%        Indels:         23
DB:                     8             Gaps:           5

US-09-610-313B-32 (1-2457) x ADN36406 (1-1457)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
Db        383 LeuAlaGluAlaMetSerGlnAla---AsnGlyThrIleLeuMetGlnArgSerAsnPhe 401

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
Db        402 LysGlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 421

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
Db        422 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 441

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
Db        442 LysAspCysThrGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProProHisLysGl 461

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
Db        461 yArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPh 481

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
Db        481 eGlyPheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuTh 501

Qy        373 TTC-----------------------------------CCCCAGATCACCCTGTGGCA 395
Db        501 rSerLeuLysSerLeuPheGlySerAspProLeuSerGlnProGlnIleThrLeuTrpGl 521

Qy        396 GCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGG 455
Db        521 nArgProLeuValSerIleArgValGlyGlyGlnIleLysGluAlaLeuLeu-------- 538

Qy        456 CGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGAT 515
Db        539 ----AspAspThrValLeuGluGluValAsnLeuProGlyLysTrpLysProLysMetIl 557

Qy        516 CGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTG 575
Db        557 eGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIleGluIleCy 577

Qy        576 CGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCG 635
Db        577 sGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyAr 597

Qy        636 CAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGT 695
Db        597 gAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrIl 617

Qy        696 GCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGA 755
Db        617 eProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrGl 637

Qy        756 GGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCAC 815
Db        637 uGluLysIleLysAlaLeuThrAlaIleCysAspGluMetGluLysGluGlyLysIleTh 657

Qy        816 CAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAG 875
Db        657 rLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSe 677

Qy        876 CACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTG 935
Db        677 rThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTr 697

Qy        936 GGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGT 995
```

```
Db        697 pGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVa 717

Qy        996 GCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACAC 1055
              ||||||||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db        717 lLeuAspValGlyAspAlaTyrPheSerValProLeuTyrGluAspPheArgLysTyrTh 737

Qy       1056 CGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGT 1115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 rAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVa 757

Qy       1116 GCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175
              |||||||||||||||||||||||||||:::|||||||||   ||||||   |||||||
Db        757 lLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetAlaLysIleLe 777

Qy       1176 GGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTA 1229
              ||||||||||||||:::||||||||||||||||||||||||||            |||||
Db        777 uGluProPheArgAlaGlnAsnProGluIleValIleTyrGlnTyrGlyAspAspLeuTy 797

Qy       1230 CGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCA 1289
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:::||
Db        797 rValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHi 817

Qy       1290 CCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCT 1349
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 sLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLe 837

Qy       1350 G------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGA 1403
              |               |||||||||||||||||||||||||||||||||:::||||||||
Db        837 uTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGl 857

Qy       1404 GAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAG 1463
              ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 uLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSe 877

Qy       1464 CCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGC 1523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 rGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAl 897

Qy       1524 CCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGA 1583
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        897 aLeuThrAspIleIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGl 917

Qy       1584 GATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGA 1643
              ||||||:::||||||||||||||||   |||||||||||||||||||||||:::||||||
Db        917 uIleLeuLysGluProValHisGlyAlaTyrTyrAspProSerLysAspLeuIleAlaGl 937

Qy       1644 GATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAA 1703
              ||||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db        937 uIleGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAs 957

Qy       1704 CCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCT 1763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 nLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLe 977

Qy       1764 GACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAA 1823
              |||||||||||||||||||||:::||||||||||||||||||||||||||   ||||||
Db        977 uThrGluAlaValGlnLysIleSerMetGluSerIleValIleTrpGlyLysIleProLy 997

Qy       1824 GTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGC 1883
              ||||||||||||||||   ||||||||||||||||||   |||||   ||||||||||||
Db        997 sPheArgLeuProIleProLysGluThrTrpGluThrArgTrpThrAlaTyrTrpGlnAl 1017

Qy       1884 CACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCA 1943
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 aThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGl 1037

Qy       1944 GCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCG 2003
              |||||||||:::||||||   |||   |||||||||||||||||||||||||||||||||
Db       1037 nLeuGluLysAspProIleAlaGlyValGluThrPheTyrValAspGlyAlaAlaAsnAr 1057

Qy       2004 CGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAG 2063
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db       1057 gGluThrLysMetGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSe 1077

Qy       2064 CCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGA 2123
              |||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db       1077 rLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleCysLeuAlaLeuGlnAs 1097

Qy       2124 CAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGC 2183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1097 pSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAl 1117

Qy       2184 CCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAA 2243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1117 aGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLy 1137

Qy       2244 GGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGAT 2303
              ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||::
Db       1137 sGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVa 1157

Qy       2304 CGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              :||||||||||||||   |||||||||||||||||||||||||||||||||||||
Db       1157 lAspLysLeuValSerAsnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1175
```

RESULT 4
AAB69289

```
ID   AAB69289 standard; protein; 1003 AA.
XX
AC   AAB69289;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 94IN476-104 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1003 AA;

Alignment Scores:
Pred. No.:              1.8e-198      Length:         1003
Score:                  3577.00       Matches:        673
Percent Similarity:     97.49%        Conservative:   26
Best Local Similarity:  93.86%        Mismatches:     14
```

```
Query Match:           78.53%      Indels:      4
DB:                    3           Gaps:        2

US-09-610-313B-32 (1-2457) x AAB69289 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerLysGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
Db         21 AlaArgAlaAsnSerProThrSerArgGluLeuGlnValGlnGlyAspAsnProArgSer 40

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGC 399
Db         41 GluAlaGlyValGluArgGlnGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg 60

Qy        400 CCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCC 459
Db         61 ProLeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla 80

Qy        460 GACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGC 519
Db         81 AspAspThrValLeuGluGluIleAlaLeuProGlyArgTrpLysProLysMetIleGly 100

Qy        520 GGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGC 579
Db        101 GlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGly 120

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
Db        121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsp 140

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
Db        141 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 160

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
Db        161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
Db        181 LysIleLysAlaLeuThrGluIleCysLysGluMetGluLysGluGlyLysIleThrLys 200

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
Db        201 IleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysArgLysAspSerThr 220

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
Db        221 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 240

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
Db        241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 260

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
Db        261 AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheGlyLysTyrThrAla 280

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
Db        281 PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 300

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
Db        301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTG 1233
Db        321 ProPheArgAlaArgAsnProLysIleValIleTyrGlnTyrMetAspAspLeuTyrVal 340

Qy       1234 GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTG 1293
Db        341 GlySerAspLeuGluIleGlyHisHisArgAlaLysIleGluGluLeuArgAlaHisLeu 360

Qy       1294 CTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG--- 1350
Db        361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 380

Qy       1351 ---CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAG 1407
Db        381 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleLysLeuProGluLys 400

Qy       1408 GAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG 1467
Db        401 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln 420

Qy       1468 ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTG 1527
Db        421 IleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440

Qy       1528 ACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATC 1587
Db        441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIle 460

Qy       1588 CTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATC 1647
Db        461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480

Qy       1648 CAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTG 1707
Db        481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500

Qy       1708 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACC 1767
Db        501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520

Qy       1768 GAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTC 1827
Db        521 GluAlaValGlnLysIleAlaIleGluSerIleValIleTrp***LysThrProLysPhe 540

Qy       1828 CGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACC 1887
Db        541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThr 560

Qy       1888 TGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTG 1947
Db        561 TrpIleProAspTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeu 580

Qy       1948 GAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAG 2007
Db        581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600

Qy       2008 ACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTG 2067
Db        601 ThrLysValGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSerLeu 620

Qy       2068 ACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGC 2127
Db        621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 640

Qy       2128 GGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAG 2187
Db        641 GlyThrGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln 660

Qy       2188 CCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAG 2247
Db        661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleAsnLysGlu 680

Qy       2248 AAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGAC 2307
Db        681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAsp 700

Qy       2308 AAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
```

```
                   :::|||||||||   ||||||||||||||||||||||||||||||||||||
Db        701 ArgLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

RESULT 5
AAB69286
ID   AAB69286 standard; protein; 999 AA.
XX
AC   AAB69286;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 96ZM651-8 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 999 AA;

Alignment Scores:
Pred. No.:              5.97e-198     Length:        999
Score:                  3568.00       Matches:       673
Percent Similarity:     97.21%        Conservative:  24
Best Local Similarity:  93.86%        Mismatches:    16
Query Match:            78.33%        Indels:        4
DB:                     3             Gaps:          2

US-09-610-313B-32 (1-2457) x AAB69286 (1-999)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AlaArgAlaAsnSerProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSer 40

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGC 399
              |||||||||   ||||||||||||:::|||||||||||||||||||||||||||||||||
Db         41 GluAlaGlyValGluArgGlnGlySerLeuAsnPheProGlnIleThrLeuTrpGlnArg 60

Qy        400 CCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCC 459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ProLeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla 80

Qy        460 GACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGC 519
                 ||||||||||||||||||::::::|||||||||||||||||||||||||||||||||
Db         81 GlyAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGly 100

Qy        520 GGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGC 579
              ||||||||||||||||:::|||||||||||||||||||||    :::|||||||||||||
Db        101 GlyIleGlyGlyPheIleGluValArgGlnTyrAspGlnIleProMetGluIleCysGly 120

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 140

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 160

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LysIleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLys 200

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThr 220

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 240

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 260

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
              ||||||||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db        261 AspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThrAla 280

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        281 PheThrIleProSerThrAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 300

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTG 1233
              ||||||||||||:::||||||:::||||||||||||||||         |||||||||
Db        321 ProPheArgAlaGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrVal 340

Qy       1234 GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTG 1293
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        341 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeu 360

Qy       1294 CTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG--- 1350
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 380

Qy       1351 ---CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAG 1407
                       |||||||||||||||||||||||||||||||:::|||   ||||||
Db        381 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuAlaGluLys 400

Qy       1408 GAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG 1467
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln 420

Qy       1468 ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTG 1527
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 IleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440

Qy       1528 ACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATC 1587
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnLysGluIle 460

Qy       1588 CTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATC 1647
              |||:::||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480

Qy       1648 CAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTG 1707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500

Qy       1708 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACC 1767
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520

Qy       1768 GAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTC 1827
              |||||||||||||||||||:::||||||||||||||||||||||   |||||||||
Db        521 GluAlaValGlnLysIleAlaLeuGluSerIleValIleTrpGlyLysIleProLysPhe 540

Qy       1828 CGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACC 1887
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThr 560

Qy       1888 TGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTG 1947
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        561 TrpIleProGluTrpGluPheValAsnThrProLeuLeuValLysLeuTrpTyrGlnLeu 580

Qy       1948 GAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAG 2007
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600

Qy       2008 ACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTG 2067
              |||||:::|||||||||||||||:::||||||||||||||||||||||||||||:::|||
Db        601 ThrLysLeuGlyLysAlaGlyTyrIleThrAspArgGlyArgGlnLysIleValThrLeu 620

Qy       2068 ACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGC 2127
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSer 640

Qy       2128 GGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAG 2187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaHis 660

Qy       2188 CCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAG 2247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlu 680

Qy       2248 AAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGAC 2307
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAsp 700

Qy       2308 AAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 LysLeuValSerLysGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

RESULT 6
AAB69287
ID   AAB69287 standard; protein; 1005 AA.
XX
AC   AAB69287;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 96ZM751-3 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1005 AA;

Alignment Scores:
Pred. No.:              1.16e-197     Length:        1005
Score:                  3563.00       Matches:       677
Percent Similarity:     96.13%        Conservative:  18
Best Local Similarity:  93.64%        Mismatches:    18
Query Match:            78.22%        Indels:        10
DB:                     3             Gaps:          3

US-09-610-313B-32 (1-2457) x AAB69287 (1-1005)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||:::||||||||||||:::|||:::|||   |||   ||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGluGlyGluAlaGlyGluLeuProSerGluGln 20

Qy        280 AACCGCGCC------------------AACAGCCCCACCAGCCGCGAGCTGCAGGTGCGC 321
              ||||||                  ||||||||||||||||||||||||||||||||||||
Db         21 ThrArgAlaAsnSerProThrSerSerAsnSerProThrSerArgGluLeuGlnValArg 40

Qy        322 GGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAG 381
              |||||||||||      ||||||||||||||||||||||||||||||      ||||||
Db         41 GlyAspAsnProCysProGluAlaGlyAlaGluArgGlnGlyThrLeuAsnCysProGln 60

Qy        382 ATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCC 441
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIleLysGluAla 80

Qy        442 CTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGG 501
              ||||||||||||||||||||||||||||||||||||||||||::::::||||||||||||
Db         81 LeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrp 100

Qy        502 AAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATC 561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIle 120

Qy        562 CTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTG 621
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        121 LeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVal 140

Qy        622 AACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGC 681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSer 160
```

```
Qy        682 CCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAG 741
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        161 ProIleGluThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGln 180

Qy        742 TGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAG 801
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLys 200

Qy        802 GAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATC 861
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIle 220

Qy        862 AAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGC 921
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArg 240

Qy        922 ACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAG 981
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLys 260

Qy        982 AAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGAC 1041
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 LysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGly 280

Qy       1042 TTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGC 1101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArg 300

Qy       1102 TACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGC 1161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSer 320

Qy       1162 ATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG--- 1218
              |||    ||||||||||||||||||||||    :::|||||||||||||||||||||
Db        321 MetIleLysIleLeuGluProPheArgThrGlnAsnProGluIleValIleTyrGlnTyr 340

Qy       1219 ---GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAG 1275
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 MetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGlu 360

Qy       1276 GAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAG 1335
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLys 380

Qy       1336 GAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCC 1389
              |||||||||||||||            |||||||||||||||||||||||||||||||||
Db        381 GluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnPro 400

Qy       1390 ATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAG 1449
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLys 420

Qy       1450 CTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTG 1509
              |||||||||   ||||||||||||   |||||||||||||||||||||||||||||||||
Db        421 LeuAsnTrp***SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeu 440

Qy       1510 CGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTG 1569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeu 460

Qy       1570 GCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAG 1629
              ||||||:::||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        461 AlaGluSerArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLys 480

Qy       1630 GACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAG 1689
              ||||||:::||||||||||||||||||||||||||||||||||||||||:::||||||
Db        481 AspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnValTyrGln 500

Qy       1690 GAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAAC 1749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsn 520

Qy       1750 GACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGG 1809
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        521 AspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrp 540

Qy       1810 GGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACC 1869
              ||||||      ||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThr 560

Qy       1870 GACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTG 1929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVal 580

Qy       1930 AAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGAC 1989
              ||||||||||||||||||||||||||||||   ||||||||||||:::||||||||||||
Db        581 LysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrTyrTyrValAsp 600

Qy       1990 GGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGG 2049
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArg 620

Qy       2050 CAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAG 2109
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 GlnLysIleValThrLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGln 640

Qy       2110 CTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTG 2169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeu 660

Qy       2170 GGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAG 2229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGlu 680

Qy       2230 CAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGC 2289
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        681 GlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGly 700

Qy       2290 GGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGAC 2349
              |||||||||:::||||||||||||||||||   |||||||||||||||||||||||||||
Db        701 GlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAsp 720

Qy       2350 GGCATCGAT 2358
              |||||||||
Db        721 GlyIleAsp 723
```

RESULT 7
AAE37601
ID   AAE37601 standard; protein; 998 AA.
XX
AC   AAE37601;
XX
DT   23-OCT-2003 (revised)
DT   27-AUG-2003 (first entry)
XX
DE   HIV-1 subtype C isolate Du151 reverse transcriptase (RT) protein.
XX
KW   Regulatory gene; accessory gene; HIV; human immunodeficiency virus;
KW   vaccine; infection; gene therapy; reverse transcriptase; RT; enzyme.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO2003037919-A2.
XX
PD   08-MAY-2003.
XX
PF   31-OCT-2002; 2002WO-IB004550.
XX
PR   31-OCT-2001; 2001ZA-00008978.
XX
PA   (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.
PA   (UYCA-) UNIV CAPE TOWN.
XX
PI   Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;
XX

```
DR   WPI; 2003-430497/40.
DR   N-PSDB; AAD29258.
XX
PT   New molecules comprising HIV-1 subtype isolate regulatory/accessory
PT   genes, useful for manufacturing a vaccine for treating or preventing HIV
PT   infection.
XX
PS   Disclosure; Page 94-97; 97pp; English.
XX
CC   The invention relates to molecules comprising HIV-1 subtype isolate
CC   regulatory/accessory genes (tat, nef and rev genes) and modifications and
CC   derivatives thereof. The invention also provides proteins encoded by such
CC   genes. Sequences of the invention are useful for manufacturing vaccines
CC   for treating or preventing human immunodeficiency virus (HIV) infections.
CC   They are also useful in gene therapy. The present sequence is HIV-1
CC   subtype C isolate Du151 reverse transcriptase (RT) protein. (Updated on
CC   23-OCT-2003 to standardise OS field)
XX
SQ   Sequence 998 AA;

Alignment Scores:
Pred. No.:              1.4e-195      Length:         998
Score:                  3527.00       Matches:        662
Percent Similarity:     96.51%        Conservative:   29
Best Local Similarity:  92.46%        Mismatches:     21
Query Match:            77.43%        Indels:         4
DB:                     6             Gaps:           2

US-09-610-313B-32 (1-2457) x AAE37601 (1-998)

Qy    223 TTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAAC 282
Db      1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20

Qy    283 CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAG 342
Db     21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu 40

Qy    343 GCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
Db     41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro 60

Qy    403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
Db     61 LeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp 80

Qy    463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
Db     81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 100

Qy    523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
Db    101 IleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLys 120

Qy    583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
Db    121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140

Qy    643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
Db    141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160

Qy    703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
Db    161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180

Qy    763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
Db    181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200

Qy    823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
Db    201 GlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220

Qy    883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
Db    221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240

Qy    943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
Db    241 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp 260

Qy   1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
Db    261 ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 280

Qy   1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
Db    281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 300

Qy   1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
Db    301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320

Qy   1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGC 1236
Db    321 PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 340

Qy   1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
Db    341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeu 360

Qy   1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------ 1350
Db    361 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380

Qy   1351 CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1410
Db    381 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400

Qy   1411 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1470
Db    401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420

Qy   1471 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1530
Db    421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440

Qy   1531 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1590
Db    441 AspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu 460

Qy   1591 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1650
Db    461 LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln 480

Qy   1651 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1710
Db    481 LysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 500

Qy   1711 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1770
Db    501 ThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThrGlu 520

Qy   1771 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1830
Db    521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540

Qy   1831 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1890
Db    541 LeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThrTrp 560

Qy   1891 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1950
Db    561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 580
```

```
Qy       1951 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2010
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        581 LysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 600

Qy       2011 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2070
              |||||||||||||||||||||||||||||||||||||||||||||||||:::|||:::
Db        601 LysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeuSer 620

Qy       2071 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2130
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640

Qy       2131 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2190
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        641 SerGluValAsnIleIleThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 660

Qy       2191 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2250
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        661 AspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluArg 680

Qy       2251 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2310
              ||||||||||||||||||||||||||||||   |||:::   |||||||||   :::||||||
Db        681 ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700

Qy       2311 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||||||||:::||||||||||||||||||   ||||||||||||||
Db        701 LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716
```

RESULT 8
AAB86169

```
ID   AAB86169 standard; protein; 3025 AA.
XX
AC   AAB86169;
XX
DT   11-SEP-2003  (revised)
DT   21-AUG-2001  (first entry)
XX
DE   HIV-1 subtype C protein fragment #1.
XX
KW   Infection; diagnosis; human; humoral immune response; antiviral;
KW   cellular immune response; vaccine; treatment; gene therapy.
XX
OS   Human immunodeficiency virus 1.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .3025
FT                   /label= Xaa
FT                   /note= "Xaa represents a stop codon"
XX
PN   DE10056747-A1.
XX
PD   31-MAY-2001.
XX
PF   16-NOV-2000; 2000DE-01056747.
XX
PR   16-NOV-1999;   99DE-01055089.
XX
PA   (SHAO/) SHAO Y.
PA   (GENE-) GENEART GMBH GES ANGEWANDTE BIOTECHNOLOG.
XX
PI   Wagner R,  Wolf H,  Shao Y,  Graf M;
XX
DR   WPI; 2001-336417/36.
DR   N-PSDB; AAH20870.
XX
PT   New nucleic acid sequences from a human immune deficiency virus
PT   intersubtype, useful for treatment, prevention and diagnosis of
PT   infection.
XX
PS   Disclosure; Fig 8A-O; 48pp; German.
XX
CC   This invention describes a novel polynucleotide isolated from human
CC   immunodeficiency virus type 1 subtype C/B' which can be used for the
CC   induction of specific humoral and cellular immune responses. (I) and
CC   polypeptides (II) encoded by them, are useful in pharmaceuticals,
CC   vaccines and diagnostic agents, particularly for treatment or prevention
CC   of human immune deficiency virus-1 (HIV-1) infections, also for rational
CC   design of test or therapeutic reagents, or gene therapy vectors.
CC   Polypeptides, especially antibodies, specifically directed against (II)
CC   are similarly useful as pharmaceutical and diagnostic agents. (I) are
CC   specific for intersubtype C/B' of HIV-1 so are useful in regions
CC   (particularly China and South-East Asia) where this subtype is prevalent.
CC   The products of the invention have antiviral activity. This sequence
CC   represents a protein encoded by the HIV-1 subtype C genome described in
CC   the method of the invention. (Updated on 11-SEP-2003 to standardise OS
CC   field)
XX
SQ   Sequence 3025 AA;

Alignment Scores:
Pred. No.:              4.55e-195    Length:        3025
Score:                  3520.00      Matches:       678
Percent Similarity:     91.14%       Conservative:  42
Best Local Similarity:  85.82%       Mismatches:    59
Query Match:            77.28%       Indels:        11
DB:                     4            Gaps:          6

US-09-610-313B-32 (1-2457) x AAB86169 (1-3025)

Qy         15 GGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAA 74
              |||||||||:::||||||         |||          |||   ||||||    :::|||
Db        416 GlyArgGlyAsnGluProAsnLysGln---CysHisThrAspAlaGluLysGlnPhe*** 434

Qy         75 GGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAA 134
                          :::         :::   ||||||   ||||||     |||    :::
Db        435 ArgLeu***LysAsnCys***MetPheGlnLeuTrpGlnGlyArgAlaHisSerGlnLys 454

Qy        135 CTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAA 194
              |||      |||    ::::::||||||||||||:::      :::          |||:::|||
Db        455 LeuGlnGlyPro***GluLysGlyLeuLeuGluMetTrpLysArgArgThrProAsnGlu 474

Qy        195 GGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAA 254
                 |||:::            |||||| ||||||||||||:::||||||   |||||||||::
Db        475 ArgLeuTyr***GluThrGlyGln-PhePheArgGluAsnLeuAlaLeuProGlnGlyAr 494

Qy        255 GGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCA 314
              :||||||||||||   |||||||||   |||||||||||||||||      ||||||||
Db        494 gAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerProThrArgGlyGluLeuGl 514

Qy        315 GGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCT 368
              ||||      |||     :::|||      |||||||||||||||   ||||||||||||::
Db        514 nValTrpGlyArgAspAsnAsnSerIleSerGluAlaGlyAlaAsnArgGlnGlyThrIl 534

Qy        369 G------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGG 422
              :         |||||||||||||||||||||||||||||||||||:::||||||:::||
Db        534 eSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGl 554

Qy        423 CGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGAT 482
              ||||||:::||||||||||||||||||||||||||||     |||||||||||||||:::::
Db        554 yGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaGlyAspThrValLeuGluAspLe 574

Qy        483 GAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGT 542
              ::::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 uAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysVa 594

Qy        543 GCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCT 602
              ||||||||||:::||||||   |||||||||||||||   ||||||||||||||||||||
Db        594 lArgGlnTyrGluGlnIleProIleGluIleCysGlyHisLysAlaIleGlyThrValLe 614

Qy        603 GATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCAC 662
              |:::|||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        614 uValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysTh 634

Qy        663 CCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGA 722
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 rLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAs 654
```

```
Qy    723 CGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCAT 782
Db    654 pGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIl 674

Qy    783 CTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAA 842
Db    674 eCysAspGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAs 694

Qy    843 CACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTT 902
Db    694 nThrProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPh 714

Qy    903 CCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCC 962
Db    714 eArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPr 734

Qy    963 CGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAG 1022
Db    734 oAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSe 754

Qy   1023 CGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAA 1082
Db    754 rIleProLeuTyrGluAspPheArgLysTyrThrAlaPheThrIleProSerArgAsnAs 774

Qy   1083 CGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCC 1142
Db    774 nGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerLe 794

Qy   1143 CAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGA 1202
Db    794 uAlaIlePheGlnSerSerMetThrLysThrLeuGluProPheArgLysGlnAsnProGl 814

Qy   1203 GATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCA 1256
Db    814 yIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGl 834

Qy   1257 GCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCC 1316
Db    834 nHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrPr 854

Qy   1317 CGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGA 1370
Db    854 oAspLysLysHis---LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAs 873

Qy   1371 CAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACAT 1430
Db    873 pLysTrpThrValGlnProThrGlnLeuProGluLysAspSerTrpThrValAsnAspIl 893

Qy   1431 CCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCG 1490
Db    893 eGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValAr 913

Qy   1491 CCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGA 1550
Db    913 gGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGl 933

Qy   1551 GGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGT 1610
Db    933 uGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyVa 953

Qy   1611 GTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTG 1670
Db    953 lTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGluGlnTr 973

Qy   1671 GACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGAT 1730
Db    973 pThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMe 993

Qy   1731 GCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCAT 1790
Db    993 tArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMe 1013

Qy   1791 GGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGAC 1850
Db   1013 tGluGlyIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluTh 1033

Qy   1851 CTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGT 1910
Db   1033 rTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVa 1053

Qy   1911 GAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGC 1970
Db   1053 lAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysAspProIleValGlyVa 1073

Qy   1971 CGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTA 2030
Db   1073 lGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTy 1093

Qy   2031 CGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGAC 2090
Db   1093 rValThrAspArgGlyArgLysLysIleValSerLeuThrGluThrThrAsnGlnLysTh 1113

Qy   2091 CGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGAC 2150
Db   1113 rGluLeuGlnAlaIleCysIleAlaLeuGlnAspSerGlySerGluValAsnIleValTh 1133

Qy   2151 CGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCT 2210
Db   1133 rAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLe 1153

Qy   2211 GGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCC 2270
Db   1153 uValAsnGlnIleIleGluGlnLeuMetLysLysGluArgValTyrLeuSerTrpValPr 1173

Qy   2271 CGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCG 2330
Db   1173 oAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleAr 1193

Qy   2331 CAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db   1193 gLysValLeuPheLeuAspGlyIleAsp 1202
```

```
RESULT 9
AAM48949
ID   AAM48949 standard; protein; 854 AA.
XX
AC   AAM48949;
XX
DT   29-AUG-2003  (revised)
DT   19-APR-2002  (first entry)
XX
DE   HIV-1 subtype C isolate Du151 pol protein.
XX
KW   HIV-1 subtype C; vaccine; HIV infection; AIDS; pol; antiviral.
XX
OS   Human immunodeficiency virus; type I.
XX
FH   Key                 Location/Qualifiers
FT   Misc-difference     58
FT                       /note= "encoded by TGA"
FT   Misc-difference     72
FT                       /note= "encoded by TAA"
FT   Misc-difference     114
FT                       /note= "encoded by TAA"
FT   Misc-difference     117
FT                       /note= "encoded by TAA"
XX
PN   WO200204494-A2.
XX
PD   17-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-IB001208.
XX
PR   07-JUL-2000; 2000US-0216995P.
PR   10-JUL-2000; 2000ZA-00003437.
```

```
PR   15-SEP-2000; 2000ZA-00004924.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
PA   (UYCA-) UNIV CAPE TOWN.
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Williamson C,  Swanstrom RI,  Morris L,  Karim SA,  Johnston RE;
XX
DR   WPI; 2002-171700/22.
DR   N-PSDB; AAL41592.
XX
PT   Selecting HIV-1 subtype C isolates, which are useful in developing
PT   vaccines against HIV infection, comprises isolating viruses with high
PT   sequence identity to a consensus sequence whose phenotype is associated
PT   with the HIV subtype.
XX
PS   Claim 19; Page 67; 69pp; English.
XX
CC   The present invention relates to a process for the selection of human
CC   immunodeficiency virus (HIV) subtype isolates for use in the development
CC   of a prophylactic and/or therapeutic pharmaceutical composition. The
CC   process involves selecting isolated virus or viruses with a high sequence
CC   identity to a consensus sequence and a phenotype which is associated with
CC   transmission for the particular HIV subtype. The composition can be used
CC   in the production of vaccines against HIV. The present sequence is the
CC   HIV-1 subtype C isolate Du151 pol protein. (Updated on 29-AUG-2003 to
CC   standardise OS field)
XX
SQ   Sequence 854 AA;

Alignment Scores:
Pred. No.:              9.24e-193     Length:        854
Score:                  3478.00       Matches:       662
Percent Similarity:     93.77%        Conservative:  30
Best Local Similarity:  89.70%        Mismatches:    31
Query Match:            76.36%        Indels:        15
DB:                     5             Gaps:          4

US-09-610-313B-32 (1-2457) x AAM48949 (1-854)

Qy     171 CGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTT------ 224
Db     118 ArgGlnGlyPheProSerHisAspValValLysArgArgProValProSerLeuHisAla 137

Qy     225 ---------------------CCGCGAGGACCTGGCC---TTCCCCCAGGGCAAGGCCCG 260
Db     138 CysArgSerThrLeuGluAspProArgValPro-SerSerPheProGlnGlyProAlaAr 157

Qy     261 CGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCG 320
Db     157 gGlnPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeuGlnValAr 177

Qy     321 CGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCA 380
Db     177 gArgAspAsnProArgSerGluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGl 197

Qy     381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db     197 nIleThrLeuTrpGlnArgProLeuValSerIleLysIleGlyGlyGlnThrArgGluAl 217

Qy     441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db     217 aLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTr 237

Qy     501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db     237 pLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIl 257

Qy     561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db     257 eLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVa 277

Qy     621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db     277 lAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSe 297

Qy     681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db     297 rProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGl 317

Qy     741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
Db     317 nTrpProLeuThrGluValLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLy 337

Qy     801 GGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCAT 860
Db     337 sGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIlePheAlaIl 357

Qy     861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920
Db     357 eLysLysGluAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysAr 377

Qy     921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAA 980
Db     377 gThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLy 397

Qy     981 GAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGA 1040
Db     397 sLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGl 417

Qy    1041 CTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCG 1100
Db     417 yPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleAr 437

Qy    1101 CTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAG 1160
Db     437 gTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSe 457

Qy    1161 CATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG-- 1218
Db     457 rMetThrLysIleLeuGluProPheArgAlaLysAsnProGluIleValIleTyrGlnTy 477

Qy    1219 ----GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGA 1274
Db     477 rMetAlaAlaLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGl 497

Qy    1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAA 1334
Db     497 uGluLeuArgGluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLy 517

Qy    1335 GGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCC 1388
Db     517 sGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnPr 537

Qy    1389 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1448
Db     537 oIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLy 557

Qy    1449 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT 1508
Db     557 sLeuAsnTrpThrSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLe 577

Qy    1509 GCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCT 1568
Db     577 uArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLe 597

Qy    1569 GGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAA 1628
Db     597 uAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLy 617

Qy    1629 GGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCA 1688
Db     617 sAspLeuIleAlaGluIleGlnLysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGl 637

Qy    1689 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAA 1748
```

```
Db        637 nGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysArgArgThrThrHisThrAs 657

Qy       1749 CGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1808
Db        657 nAspValLysGlnLeuThrGluAlaValGlnLysIleSerLeuGluSerIleValThrTr 677

Qy       1809 GGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1868
Db        677 pGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpTh 697

Qy       1869 CGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGT 1928
Db        697 rAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVa 717

Qy       1929 GAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1988
Db        717 lLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPheTyrValAs 737

Qy       1989 CGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCG 2048
Db        737 pGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyAr 757

Qy       2049 GCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCA 2108
Db        757 gGlnLysIleValThrLeuSerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGl 777

Qy       2109 GCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2168
Db        777 nLeuAlaLeuGlnAspSerGluSerGluValAsnIleValThrAspSerGlnTyrAlaLe 797

Qy       2169 GGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGA 2228
Db        797 uGlyIleIleGlnAlaGlnProAspArgSerGluSerGluLeuValAsnGlnIleIleGl 817

Qy       2229 GCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGG 2288
Db        817 uGlnLeuIleLysLysGluArgAlaTyrLeuSerTrpValProAlaHisLysGlyIleGl 837

Qy       2289 CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
Db        837 yGlyAspGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 854
```

RESULT 10
AAB69282
ID   AAB69282 standard; protein; 1000 AA.
XX
AC   AAB69282;
XX
DT   12-SEP-2003 (revised)
DT   20-APR-2001 (first entry)
XX
DE   HIV-1 non-subtype B clone 92RW009-6 pol protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH, Shaw GM, Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 15; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1000 AA;

Alignment Scores:

Pred. No.:	1.52e-192	Length:	1000
Score:	3474.50	Matches:	656
Percent Similarity:	95.40%	Conservative:	29
Best Local Similarity:	91.36%	Mismatches:	28
Query Match:	76.28%	Indels:	5
DB:	3	Gaps:	3

US-09-610-313B-32 (1-2457) x AAB69282 (1-1000)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArgLysPheSerProGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
Db         21 ThrGlyAlaAsnSerProThrSerArgGluLeuTrpAsnGlyGlyArgAspSerLeuSer 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAG 396
Db         41 SerGluThrGlyAlaGluArgGlnGlyThrPheAsnPheProGlnIleThrLeuTrpGln 60

Qy        397 CGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGC 456
Db         61 ArgProLeuValThrValLysIleGlyGlyGlnLeuArgGluAlaLeuLeuAspThrGly 80

Qy        457 GCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATC 516
Db         81 AlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIle 100

Qy        517 GGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGC 576
Db        101 GlyGlyIleGlyGlyPheIleLysValLysGlnTyrAspGlnIleLeuIleGluIleCys 120

Qy        577 GGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGC 636
Db        121 GlyLysLysAlaIleGlyThrValLeuValGlyProThrSerValAsnIleIleGlyArg 140

Qy        637 AACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTG 696
Db        141 AsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal 160

Qy        697 CCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAG 756
Db        161 ProValAlaLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 180

Qy        757 GAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACC 816
Db        181 GluLysIleLysAlaLeuArgGluIleCysThrGluMetGluLysGluGlyLysIleSer 200

Qy        817 AAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGC 876
Db        201 LysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 220

Qy        877 ACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936
Db        221 ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp 240
```

```
Qy        937 GAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTG 996
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVal 260

Qy        997 CTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACC 1056
              ||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        261 LeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThr 280

Qy       1057 GCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTG 1116
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 300

Qy       1117 CTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTG 1176
              ||||||||||||||||||||||||:::|||||||||:::|||||||||||||||||||||
Db        301 LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAsnSerMetThrLysIleLeu 320

Qy       1177 GAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTAC 1230
              |||||||||||||:::|||   ||||||||||||||||||            ||||||
Db        321 GluProPheArgAlaGlnAsnGlnGluIleValIleTyrGlnTyrMetAspAspLeuTyr 340

Qy       1231 GTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCAC 1290
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||
Db        341 ValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHis 360

Qy       1291 CTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG 1350
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 380

Qy       1351 ------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAG 1404
                          |||||||||||||||||||||||||||||||||:::||||||||||
Db        381 TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGlu 400

Qy       1405 AAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGC 1464
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 420

Qy       1465 CAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCC 1524
              |||||||||||||||:::|||||||||||||||||||||||||||||||   ||||||||
Db        421 GlnIleTyrProGlyValLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAla 440

Qy       1525 CTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAG 1584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 LeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 460

Qy       1585 ATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAG 1644
              ||||||:::|||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        461 IleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGlu 480

Qy       1645 ATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAAC 1704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 IleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 500

Qy       1705 CTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTG 1764
              |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        501 LeuLysThrGlyLysTyrAlaLysArgArgThrAlaHisThrAsnAspValLysGlnLeu 520

Qy       1765 ACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAG 1824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 ThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrpGlyLysThrProLys 540

Qy       1825 TTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCC 1884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAla 560

Qy       1885 ACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAG 1944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 580

Qy       1945 CTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGC 2004
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        581 LeuGluLysGluProIleLeuGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg 600

Qy       2005 GAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGC 2064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSer 620

Qy       2065 CTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGAC 2124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 LeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAsp 640

Qy       2125 AGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCC 2184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 SerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 660

Qy       2185 CAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAG 2244
              |||||||||   ||||||||||||   |||||||||||||||||||||||||||||||||
Db        661 GlnProAspSerSerGluSerGluAlaValAsnGlnIleIleGluGlnLeuIleLysLys 680

Qy       2245 GAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATC 2304
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        681 GluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 700

Qy       2305 GACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||||||||||||||   |||||||||:::||||||||||||||||||||||||
Db        701 AspLysLeuValSerSerGlyIleArgArgValLeuPheLeuAspGlyIleAsp 718
```

```
RESULT 11
AAW72993
ID   AAW72993 standard; protein; 1002 AA.
XX
AC   AAW72993;
XX
DT   25-MAR-2003  (revised)
DT   15-FEB-1999  (first entry)
XX
DE   HIV isolate LAV.MAL pol protein.
XX
KW   LAV.MAL; HIV; human immunodeficiency virus; diagnosis; vaccine; AIDS;
KW   pol.
XX
OS   Human T-lymphotropic virus.
XX
PN   US5824482-A.
XX
PD   20-OCT-1998.
XX
PF   06-JUN-1995;   95US-00471474.
XX
PR   23-JUN-1986;   86FR-00040138.
PR   13-APR-1987;   87US-00038330.
PR   19-FEB-1991;   91US-00656797.
PR   10-DEC-1992;   92US-00988530.
PR   18-NOV-1993;   93US-00154397.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Wain-Hobson S,  Sonigo P,  Alizon M,  Montagnier L;
XX
DR   WPI; 1998-582548/49.
DR   N-PSDB; AAV63467.
XX
PT   Human immunodeficiency virus isolate LAV(MAL) - and method for detecting
PT   anti-HIV antibodies.
XX
PS   Disclosure; Fig 7B-E; 47pp; English.
XX
CC   This is the amino acid sequence of the pol protein of lymphadenopathy
CC   associated virus LAV.MAL (CNCM I-641), a new virus isolate from Zaire
CC   that is responsible for diseases clinically related to AIDS. The sequence
CC   was deduced from an open reading frame (ORF) of the LAV.MAL genome (see
CC   AAV63467). 7 ORFs (see AAW72992-98) were identified. Specific peptides of
CC   the envelope glycoprotein can be used as antigens in a claimed method for
CC   the in vitro detection of an antibody directed against LAV. The method is
CC   useful for the diagnosis of AIDS or pre-AIDS, or to detect antibodies in
CC   patients, asymptomatic carriers and in blood-related products. LAV viral
```

```
CC   antigens are also useful in vaccines. (Updated on 25-MAR-2003 to correct
CC   PR field.)
XX
SQ   Sequence 1002 AA;

Alignment Scores:
Pred. No.:                1.68e-191     Length:         1002
Score:                    3456.50       Matches:        649
Percent Similarity:       95.56%        Conservative:   39
Best Local Similarity:    90.14%        Mismatches:     25
Query Match:              75.88%        Indels:         7
DB:                       2             Gaps:           4

US-09-610-313B-32 (1-2457) x AAW72993 (1-1002)

Qy     220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db       1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy     280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
Db      21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40

Qy     337 AGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACCCTG 390
Db      41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPheSerPheProGlnIleThrLeu 60

Qy     391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
Db      61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy     451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
Db      81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy     511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
Db     101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy     571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
Db     121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140

Qy     631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
Db     141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy     691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
Db     161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180

Qy     751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
Db     181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200

Qy     811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
Db     201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy     871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
Db     221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy     931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
Db     241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy     991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
Db     261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy    1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
Db     281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy    1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
Db     301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy    1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
Db     321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy    1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
Db     341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy    1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
Db     361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy    1345 TTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1398
Db     381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400

Qy    1399 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1458
Db     401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy    1459 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1518
Db     421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440

Qy    1519 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1578
Db     441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460

Qy    1579 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1638
Db     461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy    1639 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1698
Db     481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500

Qy    1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1758
Db     501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520

Qy    1759 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1818
Db     521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540

Qy    1819 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878
Db     541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy    1879 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1938
Db     561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy    1939 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 1998
Db     581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy    1999 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2058
Db     601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy    2059 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118
Db     621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy    2119 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2178
Db     641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660
```

```
Qy       2179 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2238
              |||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

Qy       2239 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2298
              :::|||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2299 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||:::||||||||||||||   |||||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 12
AAO30963
ID   AAO30963 standard; protein; 1003 AA.
XX
AC   AAO30963;
XX
DT   06-OCT-2003  (first entry)
XX
DE   HIV pol protein.
XX
KW   Human immunodeficiency virus; HIV; pol; IV9; immunotherapy; infection;
KW   epitope; vaccine; enzyme.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003051285-A2.
XX
PD   26-JUN-2003.
XX
PF   29-OCT-2002; 2002WO-US034620.
XX
PR   29-OCT-2001; 2001US-0345116P.
XX
PA   (GENZ ) GENZYME CORP.
PA   (GEHO ) GEN HOSPITAL CORP DBA MASSACHUSETTS GEN.
XX
PI   Nicolette CA,  Walker BD;
XX
DR   WPI; 2003-541603/51.
XX
PT   Novel synthetic therapeutic peptide compounds useful for inducing an
PT   immune response in a mammal, specifically cross-react with HIV IV9
PT   epitopes.
XX
PS   Claim 8; Page 65-67; 67pp; English.
XX
CC   The invention relates to synthetic therapeutic peptide compounds that
CC   specifically cross-react with HIV (human immunodeficiency virus) pol
CC   protein IV9 epitopes. The invention is useful for treating HIV infection.
CC   The invention is also used in immunotherapy and as vaccines. The present
CC   sequence is HIV pol protein
XX
SQ   Sequence 1003 AA;

Alignment Scores:
Pred. No.:              2.34e-191     Length:         1003
Score:                  3454.00       Matches:        651
Percent Similarity:     95.01%        Conservative:   34
Best Local Similarity:  90.29%        Mismatches:     28
Query Match:            75.83%        Indels:         8
DB:                     6             Gaps:           4

US-09-610-313B-32 (1-2457) x AAO30963 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||   |||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
              |||||||||||||||||   |||||||||||||||      |||   :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
              ||||||||||||||:::||||||||||||:::      ||||||||||||:::|||
Db         41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              ||||||||||||||||||:::||||||:::|||||||||:::|||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              ||||||||||||   |||||||||||||||||||||:::|||||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||||:::|||||||||:::||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              |||||||||||||||||||||||||||||||      ||||||   ||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              ||||||||||||||||||   :::||||||:::||||||||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
              |||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::|||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
              |||||||||      ||||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
```

```
Qy       1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
                |||||||||||::::::::::||||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1636 GTGGCCGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
              :::||||||||||||||||||||      |||||||||||||||||||||||||||||||
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
              |||||||||||||||||||||||||||||:::||||||      ||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy       1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
              |||||||||||||||||||||||||||        |||||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy       1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||:::|||
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
              ||||||||||||||||||:::|||||||||||||||||||||:::|||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
              :::|||:::||||||:::|||||||||||||||||||||||||||||||      |||||
Db        621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy       2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
              ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
              ||||||:::||||||||||||||||      |||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2356 GAT 2358
              |||
Db        721 Asp 721
```

RESULT 13
ADN36413
ID   ADN36413 standard; protein; 1003 AA.
XX
AC   ADN36413;
XX
DT   15-JUL-2004 (first entry)
XX
DE   HIV protein for anti-HIV vaccine.
XX
KW   anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO2004035006-A2.
XX
PD   29-APR-2004.
XX
PF   17-OCT-2003; 2003WO-US033112.
XX
PR   18-OCT-2002; 2002US-0419465P.
XX
PA   (AARO-) AARON DIAMOND AIDS RES CENT.
XX
PI   Huang Y, Ho DD, Chen Z;
XX
DR   WPI; 2004-348328/32.
DR   N-PSDB; ADN36412.
XX
PT   Nucleic acid vector comprising at least one HIV sequence operably linked
PT   to a promoter and encoding a protein that does not assemble into viral
PT   particles, useful in immunizing a subject against HIV infection.
XX
PS   Disclosure; SEQ ID NO 27; 166pp; English.
XX
CC   The invention relates to a nucleic acid vector comprising at least one
CC   HIV sequence operably linked to a promoter and encoding a protein that
CC   does not assemble into viral particles. The nucleic acid vector is useful
CC   in immunizing a subject against HIV infection. This sequence corresponds
CC   to a nucleic acid used in the invention.
XX
SQ   Sequence 1003 AA;

```
Alignment Scores:
Pred. No.:               2.34e-191      Length:         1003
Score:                   3454.00        Matches:        651
Percent Similarity:      95.01%         Conservative:   34
Best Local Similarity:   90.29%         Mismatches:     28
Query Match:             75.83%         Indels:         8
DB:                      8              Gaps:           4

US-09-610-313B-32 (1-2457) x ADN36413 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
              |||||||||||||||||   ||||||||||||||      |||   :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
              |||||||||||||||:::|||||||||||||:::      ||||||||||||:::|||
Db         41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||||:::||||||:::|||||||||:::|||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
```

```
Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db    321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy   1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db    341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy   1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db    361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy   1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
Db    381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy   1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db    401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy   1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db    421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy   1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db    441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy   1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db    461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy   1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db    481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy   1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db    501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy   1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db    521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy   1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db    541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy   1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db    561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy   1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db    581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy   1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db    601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy   2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db    621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy   2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
Db    641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy   2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
Db    661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy   2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
Db    681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy   2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
Db    701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy   2356 GAT 2358
Db    721 Asp 721
```

```
RESULT 14
AAP81861
ID   AAP81861 standard; protein; 1002 AA.
XX
AC   AAP81861;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   16-DEC-1990  (first entry)
XX
DE   Sequence encoded by LAV MA L POL gene.
XX
KW   HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.
XX
OS   Lymphadenopathy-associated virus.
XX
PN   WO8707906-A.
XX
PD   30-DEC-1987.
XX
PF   22-JUN-1987;   87WO-EP000326.
XX
```

```
PR   23-JUN-1986;   86EP-00401380.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Alizon M,  Sonigo P,  Wainhobson S,  Montagnier L,  Alizon M;
PI   Sonigo P,  Wainhobson S,  Montagnier L;
XX
DR   WPI; 1988-014396/02.
DR   N-PSDB; AAN80437.
XX
PT   New variants of lymphadenopathy associated virus (LAV) - used for prodn.
PT   of DNA, antigens and antibodies used in diagnosis of AIDS and pre-AIDS.
XX
PS   Claim 8; Fig 8A-8I; 72pp; English.
XX
CC   LAV EL I (AAN80436) and LAV MA L (AAN80437) were isolated from the
CC   peripheral blood lymphocytes of patients. Different AIDS virus isolates
CC   concerned are designated by 3 letters of the patients name. Stable probes
CC   including the DNA sequences can be used for detection of the new LAV
CC   viruses or related viruses or DNA proviruses in eg. biological samples.
CC   The proteins or peptides can be used for detection of antibodies induced
CC   in vivo and present in biological fluids. The DNA can also be used for
CC   the expression of LAV viral antigens for the prodn. of a vaccine against
CC   LAV. The polypeptides can also be used for the prodn. of antibodies for
CC   the detection of proteins related to the LAV viruses, partic. for
CC   diagnosis of AIDS or pre-AIDS. (Updated on 25-MAR-2003 to correct PI
CC   field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 1002 AA;

Alignment Scores:
Pred. No.:              3.26e-191     Length:        1002
Score:                  3451.50       Matches:       648
Percent Similarity:     95.42%        Conservative:  39
Best Local Similarity:  90.00%        Mismatches:    26
Query Match:            75.77%        Indels:        7
DB:                     1             Gaps:          4

US-09-610-313B-32 (1-2457) x AAP81861 (1-1002)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
                 ||||||||||||||||||||||||||||:::|||      ||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACCCTG 390
              ||||||   ||||||||||||||||||       :::   :::||||||||||||||||
Db         41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPheSerPheProGlnIleThrLeu 60

Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||:::|||:::::::::|||||||||||||:::|||||||||||||||||||
Db         61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              ||||||||||||||||||||||||||||||::::::||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||||||||||||||||||:::|||:::||||||||||||||||||||||||||
Db        121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              |||||||||||||||||||||||||||   ||||||::::::||||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
              ||||||||||||||||    :::||||||||||||||||||||||||
Db        321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy       1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
              ||||||||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy       1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
              :::||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy       1345 TTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1398
              ||||||      ||||||||||||||||||||||||||||||||||||||||:::|||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400

Qy       1399 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1458
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy       1459 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1518
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440

Qy       1519 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1578
              |||||||||||||||||||||||||||||||   ||||||||||||||||   ||||||
Db        441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuProGluAsn 460

Qy       1579 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1638
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1639 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1698
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||   :::
Db        481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500

Qy       1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1758
              ||||||||||||||||||||||||||||:::::::::::|||||||||||||||||||||
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520

Qy       1759 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1818
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
```

```
Db        521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540

Qy       1819 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878
              ||||||||||||||||||||||||||||||||||||||    ||||||||:::||||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560

Qy       1879 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1938
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1939 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 1998
              ||||||||||||     |||||||||:::||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       1999 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2058
              |||||||||||||||     ||||||||||||||||||||||||||||||||||||:::
Db        601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2059 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118
              |||||||||||||||||||||||||||||||||||||||||||||||    |||||||||
Db        621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640

Qy       2119 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2179 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2238
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

Qy       2239 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2298
              :::|||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2299 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||:::||||||||||||||||    ||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

```
RESULT 15
ABR55489
ID   ABR55489 standard; protein; 1003 AA.
XX
AC   ABR55489;
XX
DT   11-AUG-2003  (first entry)
XX
DE   Amino acid sequence of a HIV pol protein.
XX
KW   Polymorphic marker; host response; HIV; AIDS related virus; vaccine;
KW   hepatitis related virus; HCV; HBV; drug resistance; pol.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003035097-A1.
XX
PD   01-MAY-2003.
XX
PF   23-OCT-2002; 2002WO-AU001450.
XX
PR   23-OCT-2001; 2001AU-00008425.
XX
PA   (EPIP-) EPIPOP PTY LTD.
XX
PI   Mallal S;
XX
DR   WPI; 2003-449231/42.
XX
PT   Determining the influence of variation in host genes on the selection of
PT   microorganisms with protein substitutions, comprises typing individuals
PT   of a cohort infected with a microorganism for an intrinsic polymorphic
PT   marker.
XX
PS   Claim 22; Page 88-89; 157pp; English.
XX
CC   The specification describes a method of determining the influence of
CC   variation in host genes on selection of microorganisms with protein
CC   substitutions. The method comprises typing all individuals of a
CC   population of patients infected with a microorganism for at least one
CC   selected intrinsic polymorphic marker involved in the host response to
CC   the presence of the microorganism. The method is useful for examining
CC   selective pressures confronting a wide range of organisms that exhibit
CC   pathogenic traits in a host, such as bacteria, fungi, mycobacterium,
CC   viruses and virus-like particles; for examining microorganisms that have
CC   adapted to evolve rapidly, including HIV and AIDS related viruses and the
CC   hepatitis related viruses such as HCV and HBV. The method is useful for
CC   designing a vaccine to prevent or delay the emergence of drug resistance
CC   in patients treated with a particular drug specific for a microorganism
CC   where the drug affects the replication of the microorganism at the
CC   nucleotide or amino acid level. The present sequence represents a HIV pol
CC   protein, which is expected to provide optimal cytotoxic T lymphocyte
CC   (CTL) induced therapeutic protection to the cohort examined in that study
XX
SQ   Sequence 1003 AA;

Alignment Scores:
Pred. No.:              3.49e-191      Length:         1003
Score:                  3451.00        Matches:        650
Percent Similarity:     95.15%         Conservative:   36
Best Local Similarity:  90.15%         Mismatches:     27
Query Match:            75.76%         Indels:         8
DB:                     6              Gaps:           4

US-09-610-313B-32 (1-2457) x ABR55489 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||:::||||||||||||||||||||||||||||||    |||||||||
Db          1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGAC------AACCCC 333
                 ||||||||||||||||   ||||||||||||||   |||:::      |||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGluAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||||:::|||||||||||:::      :::||||||||||||||
Db         41 ThrSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              ||||||||||||||||||||||:::||||||:::||||||:::||||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||||:::|||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleIleIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              ||||||||||||    ||||||||||||||||||||:::|||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||       ||||||   ||||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy     868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db     221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy     928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db     241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy     988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db     261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy    1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db     281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy    1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db     301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy    1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db     321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy    1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db     341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy    1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db     361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy    1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
Db     381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy    1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db     401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy    1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db     421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy    1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db     441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy    1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db     461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy    1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db     481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy    1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db     501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy    1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db     521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy    1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db     541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560

Qy    1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db     561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy    1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db     581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy    1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db     601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy    2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db     621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy    2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
Db     641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy    2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
Db     661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy    2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
Db     681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy    2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
Db     701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy    2356 GAT 2358
Db     721 Asp 721

Search completed: June  2, 2005, 04:43:53
Job time : 233.424 secs
```

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:          June  2, 2005, 02:53:39 ; Search time 40.7339 Seconds
                                           (without alignments)
                                           11607.276 Million cell updates/sec

Title:           US-09-610-313B-32
Perfect score:   4555
Sequence:        1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2457

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:      566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155137_15159/app_query.fasta_1.7893
-DB=PIR_79 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_189_@runat_31052005_155137_15159 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       PIR_79:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3462.5	76.0	1002	2	S54378	pol polyprotein -
2	3454.5	75.8	1002	1	GNLJND	HIV-1 retropepsin
3	3449	75.7	1003	1	GNVWLV	HIV-1 retropepsin
4	3432	75.3	1015	1	GNVWH3	HIV-1 retropepsin
5	3410	74.9	1003	2	T09440	pol polyprotein -
6	3407	74.8	1003	1	B44001	HIV-1 retropepsin
7	3406	74.8	1003	1	GNVWA2	HIV-1 retropepsin
8	3396	74.6	1012	1	GNVWVL	HIV-1 retropepsin
9	3168	69.5	1027	1	GNLJSI	HIV-1 retropepsin
10	3064	67.3	912	2	S33980	pol polyprotein -
11	3051	67.0	902	2	T01668	pol polyprotein -
12	2728	59.9	559	2	B47175	reverse transcript
13	2708	59.5	559	2	A47175	reverse transcript
14	2307	50.6	1039	2	S46347	pol polyprotein -
15	2170	47.6	1055	2	S53092	pol polyprotein -
16	2166.5	47.6	1055	1	GNLJST	HIV-1 retropepsin
17	2158	47.4	1036	1	GNLJG2	HIV-1 retropepsin
18	2157.5	47.4	1056	1	GNLJG3	HIV-1 retropepsin
19	2149.5	47.2	1035	1	GNLJGG	HIV-1 retropepsin
20	2149	47.2	1034	1	GNLJCA	HIV-1 retropepsin
21	2143.5	47.1	1019	2	T11560	pol polyprotein -
22	2130.5	46.8	1009	2	S28081	pol polyprotein -
23	2111.5	46.4	1032	2	S12153	pol polyprotein -
24	2034	44.7	1058	2	S08436	pol polyprotein -
25	2015	44.2	1061	1	GNLJG4	HIV-1 retropepsin
26	1980.5	43.5	1054	1	GNLJG5	HIV-1 retropepsin
27	1925.5	42.3	656	2	S30484	pol polyprotein -
28	1896.5	41.6	656	2	S30483	pol polyprotein -
29	1350	29.6	1124	2	S23820	pol polyprotein -
30	1347	29.6	1124	1	GNLJFP	HIV-1 retropepsin
31	1305	28.6	1124	2	B45557	HIV-1 retropepsin
32	1122.5	24.6	1109	1	B45345	HIV-1 retropepsin
33	1106.5	24.3	1087	2	JQ1162	Pol protein - Maed
34	1102.5	24.2	1086	1	B46335	HIV-1 retropepsin
35	1100	24.1	1146	1	GNLJ22	HIV-1 retropepsin
36	1100	24.1	1146	1	GNLJEW	HIV-1 retropepsin
37	1095	24.0	1101	1	B45390	HIV-1 retropepsin
38	1091	24.0	1101	1	GNLJVS	HIV-1 retropepsin
39	1080.5	23.7	1145	1	GNLJEV	HIV-1 retropepsin
40	1071	23.5	219	2	S32152	RNA-directed DNA p
41	1067	23.4	219	2	S32118	RNA-directed DNA p
42	1067	23.4	219	2	S32139	RNA-directed DNA p
43	1065	23.4	219	2	S32089	RNA-directed DNA p
44	1065	23.4	219	2	S32160	RNA-directed DNA p
45	1064	23.4	219	2	S32133	RNA-directed DNA p

ALIGNMENTS

```
RESULT 1
S54378
pol polyprotein - human immunodeficiency virus type 1 (fragment)
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C;Accession: S54378
R;Theodore, T.; Buckler-White, A.J.
submitted to the EMBL Data Library, July 1989
A;Reference number: S54377
A;Accession: S54378
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1-1002 <THE>
A;Cross-references: UNIPROT:P12499; EMBL:M22639; NID:g329377; PIDN:AAA45366.1; PID:g3293
C;Superfamily: pol polyprotein
C;Keywords: polyprotein

Alignment Scores:
Pred. No.:               3.6e-154       Length:        1002
Score:                   3462.50        Matches:       651
Percent Similarity:      95.14%         Conservative:  34
Best Local Similarity:   90.42%         Mismatches:    28
Query Match:             76.02%         Indels:        7
DB:                      2              Gaps:          4

US-09-610-313B-32 (1-2457) x S54378 (1-1002)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||||||||||||||||||||||  |||   ||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC---GACAACCCCCGC 336
                 ||||||||||||||||||||||||||||:::|||   |||   ||||||||
Db         21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTC------CCCCAGATCACCCTG 390
              |||||   |||||||||||||||||||::::::|||      ||||||||||||||
Db         41 SerGluThrGlyAlaGluArgGlnGlyThrValSerPheAsnCysProGlnIleThrLeu 60
```

```
Qy        391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
              ||||||||||||||||||:::||||||:::|||||||||:::||||||||||||||||||
Db         61 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy        451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db         81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||   |||||||||||||||||||||:::||||||||||||||||||||||||
Db        121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              |||||||||:::|||||||||:::||||||||||||||||||||||||||||||||||||
Db        141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              |||||||||||||||||||||||||||   ||||||   |||||||||||||||||||||
Db        181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||:::::::::|||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        201 IleSerArgValGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLys 280

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
              ||||||||||||||||||   :::||||||||||||||||||||||||
Db        321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy       1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy       1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy       1345 TTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1398
              ||||||            |||||||||||||||||||||||||||   |||:::|||
Db        381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400

Qy       1399 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy       1459 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1518
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr 440

Qy       1519 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1578
              ||||||||||||:::::::::|||||||||||||||||||||||||||||||||||||||
Db        441 LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460

Qy       1579 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1638
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||:::
Db        461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy       1639 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1698
              ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy       1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1758
              |||||||||||||||||||||||||||:::||||||   |||||||||||||||||||||
Db        501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520

Qy       1759 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1818
              ||||||   |||   ||||||||||||:::   |||||||||||||||||||||||||||
Db        521 GlnLeuAlaGluValValGlnLysIleSerThrGluSerIleValIleTrpGlyLysThr 540

Qy       1819 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878
              ||||||||||||||||||||||||||||||||||||||||||||||||   :::||||||
Db        541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpValGluTyrTrp 560

Qy       1879 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1938
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy       1939 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 1998
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy       1999 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2058
              |||||||||||||||:::|||||||||||||||||||||||||||||||||||||||:::
Db        601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy       2059 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118
              |||      |||:::|||||||||||||||||||||||||||||||||   |||||||||
Db        621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640

Qy       2119 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2178
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2179 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2238
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680

Qy       2239 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2298
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2299 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||:::|||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 2
GNLJND
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate NDK)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: JQ0067
R;Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.,
Gene 81, 275-284, 1989
A;Title: Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immuno
A;Reference number: JQ0065; MUID:90034200; PMID:2806917
A;Accession: JQ0067

```
A;Molecule type: DNA
A;Residues: 1-1002 <SPI>
A;Cross-references: UNIPROT:P18802; GB:M27323; NID:g328154; PIDN:AAA44869.1; PID:g328158
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nuclec
F;56-154/Product: retropepsin #status predicted <RTP>
F;80/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:
Pred. No.:              8.49e-154    Length:        1002
Score:                  3454.50      Matches:       651
Percent Similarity:     95.14%       Conservative:  34
Best Local Similarity:  90.42%       Mismatches:    28
Query Match:            75.84%       Indels:        7
DB:                     1            Gaps:          4

US-09-610-313B-32 (1-2457) x GNLJND (1-1002)

Qy   220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db     1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaGlyGluPheSerSerGluGln 20

Qy   280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGACAACCCCCGC 336
Db    21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspAsnProLeu 40

Qy   337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTG 390
Db    41 SerGluThrGlyAlaGluArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60

Qy   391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
Db    61 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80

Qy   451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
Db    81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100

Qy   511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
Db   101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120

Qy   571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
Db   121 IleCysGlyTyrLysAlaMetGlyThrValLeuValGlyProThrProValAsnIleIle 140

Qy   631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
Db   141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160

Qy   691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
Db   161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180

Qy   751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
Db   181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200

Qy   811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
Db   201 IleSerArgIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220

Qy   871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
Db   221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240

Qy   931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
Db   241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 260

Qy   991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
Db   261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280

Qy  1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
Db   281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300

Qy  1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
Db   301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320

Qy  1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
Db   321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340

Qy  1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
Db   341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360

Qy  1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
Db   361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380

Qy  1345 TTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1398
Db   381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleAsnLeu 400

Qy  1399 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1458
Db   401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

Qy  1459 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1518
Db   421 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 440

Qy  1519 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1578
Db   441 LysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460

Qy  1579 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1638
Db   461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480

Qy  1639 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1698
Db   481 AlaGluLeuGlnLysGlnGlyAspGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500

Qy  1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1758
Db   501 LysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspValLys 520

Qy  1759 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1818
Db   521 GlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 540

Qy  1819 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878
Db   541 ProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpIleGluTyrTrp 560

Qy  1879 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1938
Db   561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580

Qy  1939 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 1998
Db   581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600

Qy  1999 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2058
Db   601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620

Qy  2059 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118
```

```
Db        621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640

Qy       2119 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2178
Db        641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy       2179 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2238
Db        661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680

Qy       2239 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2298
Db        681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700

Qy       2299 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db        701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

RESULT 3
GNVWLV
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LAV-1a)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 03-Jun-2002
C;Accession: A03966
R;Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.
Cell 40, 9-17, 1985
A;Title: Nucleotide sequence of the AIDS virus, LAV.
A;Reference number: A90866; MUID:85099333; PMID:2981635
A;Accession: A03966
A;Molecule type: DNA
A;Residues: 1-1003 <WAI>
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nuclec
F;57-155/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status experimental

Alignment Scores:

Pred. No.:	1.53e-153	Length:	1003
Score:	3449.00	Matches:	651
Percent Similarity:	94.87%	Conservative:	33
Best Local Similarity:	90.29%	Mismatches:	29
Query Match:	75.72%	Indels:	8
DB:	1	Gaps:	4

US-09-610-313B-32 (1-2457) x GNVWLV (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db        361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db        421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
```

```
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520

Qy       1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db        521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy       1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db        601 AlaSerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db        621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2356 GAT 2358
Db        721 Asp 721
```

RESULT 4
GNVWH3
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate HTLV-III
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C;Accession: A03965
R;Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K.J.; Starcich, B.; Josephs, S.F.; Dora
nberger, J.A.; Papas, T.S.; Ghrayeb, J.; Chang, N.T.; Gallo, R.C.; Wong-Staal, F.
Nature 313, 277-284, 1985
A;Title: Complete nucleotide sequence of the AIDS virus, HTLV-III.
A;Reference number: A93353; MUID:85111123; PMID:2578615
A;Accession: A03965
A;Molecule type: DNA
A;Residues: 1-1015 <RAT>
A;Cross-references: UNIPROT:P03366; GB:M15654; GB:K02008; GB:K02009; GB:K02010; NID:g326
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;69-167/Product: retropepsin #status predicted <RTP>
F;93/Active site: Asp (shared with dimeric partner) #status experimental

```
Alignment Scores:
Pred. No.:              9.52e-153     Length:         1015
Score:                  3432.00       Matches:        650
Percent Similarity:     93.18%        Conservative:   33
Best Local Similarity:  88.68%        Mismatches:     30
Query Match:            75.35%        Indels:         20
DB:                     1             Gaps:           5
```

US-09-610-313B-32 (1-2457) x GNVWH3 (1-1015)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        268 ------------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db         21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy        304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db         41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy        358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db         61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy        412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db         81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy        472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db        101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy        532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db        121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy        592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db        141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy        652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db        161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy        712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db        181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy        772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db        201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
```

```
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy       1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAG 1359
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy       1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy       1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
Db        641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 5
T09440
pol polyprotein - human immunodeficiency virus type 1 (strain JRFL) (fragment)
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T09440
R;Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie,
submitted to the EMBL Data Library, July 1996
A;Reference number: Z16673
A;Accession: T09440
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1003 <PAN>
A;Cross-references: UNIPROT:Q75755; EMBL:U63632; NID:g1465777; PID:g1465779
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein

```
Alignment Scores:
Pred. No.:              1.01e-151     Length:        1003
Score:                  3410.00       Matches:       642
Percent Similarity:     94.45%        Conservative:  39
Best Local Similarity:  89.04%        Mismatches:    32
Query Match:            74.86%        Indels:        8
DB:                     2             Gaps:          4

US-09-610-313B-32 (1-2457) x T09440 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db         21 ThrArgAlaAsnSerProThrArgLysGluLeuGlnValTrpGlyArgAspSerAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGC------ACCCTGAACTTCCCCCAGATCACC 387
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyAsnValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db         81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
```

```
Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||:::||||||||:::||||||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||         ||||||   ||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysArgLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              ||||||||||||||||||     :::||||||:::|||:::||||||||
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
              |||||||||            |||||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
              ||||||||||||||||||     ||||||||||||:::||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
                 ||||||||||||::::::::||||||||||||||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
              :::||||||:::||||||||||||         ||||||||||||||||||||||||||||
Db        481 IleAlaGluLeuGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
              ||||||   ||||||||||||||||||||:::   |||   ||||||||||||||||||
Db        501 PheLysIleLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520

Qy       1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
                    |||||||:::||||||||||||||||||||||||||||||||||||:::|||
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
              |||||||||||||||||:::|||||||||||||||||||||||||:::||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
              :::|||||||||||:::|||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
              ||||||||||||||      ||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
              |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
              ||||||||||||||||||||||||:::||||||||||||||||||:::||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
              ||||||:::||||||||||||||||      |||||||||||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2356 GAT 2358
              |||
Db        721 Asp 721
```

RESULT 6
B44001
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (strain YU-2)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: B44001
R;Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.
J. Virol. 66, 6587-6600, 1992
A;Title: Complete nucleotide sequence, genome organization, and biological properties of
A;Reference number: A44001; MUID:93021387; PMID:1404605
A;Accession: B44001
A;Molecule type: DNA
A;Residues: 1-1003 <LIY>
A;Cross-references: UNIPROT:P35963; GB:M93258
C;Comment: This protein is synthesized as a gag-pol polyprotein.
C;Comment: Specific enzymatic cleavages may yield mature proteins including proteinase,
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;57-155/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status predicted

```
Alignment Scores:
Pred. No.:                1.39e-151     Length:        1003
Score:                    3407.00       Matches:       643
Percent Similarity:       94.31%        Conservative:  37
Best Local Similarity:    89.18%        Mismatches:    33
Query Match:              74.80%        Indels:        8
DB:                       1             Gaps:          4

US-09-610-313B-32 (1-2457) x B44001 (1-1003)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db      1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy    280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db     21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40

Qy    334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db     41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60

Qy    388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db     61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy    448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db     81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100

Qy    508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db    101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuHisGluAspPheArg 280

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyThrArgTyrGln 300

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db    321 ThrIleLeuGluProPheArgLysGlnAsnProAspLeuValIleTyrGlnTyrMetAsp 340

Qy   1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db    341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy   1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db    361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy   1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
Db    381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy   1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db    401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy   1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db    421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy   1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db    441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy   1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db    461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy   1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db    481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy   1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db    501 PheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520

Qy   1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db    521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540

Qy   1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db    541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy   1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db    561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy   1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db    581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600

Qy   1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db    601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnLysGlyArgGlnLys 620

Qy   2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db    621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy   2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
Db    641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy   2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
Db    661 IleGlnAlaGlnProAspArgSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
```

```
Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2356 GAT 2358
Db        721 Asp 721
```

RESULT 7
GNVWA2
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate ARV-2)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C;Accession: A03968
R;Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Sh
Science 227, 484-492, 1985
A;Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).
A;Reference number: A04003; MUID:85090453; PMID:2578227
A;Accession: A03968
A;Molecule type: DNA
A;Residues: 1-1003 <SAN>
A;Cross-references: UNIPROT:P03369; GB:K02007; NID:g328658; PIDN:AAB59876.1; PID:g328662
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;57-147/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status predicted

```
Alignment Scores:
Pred. No.:              1.55e-151     Length:         1003
Score:                  3406.00       Matches:        642
Percent Similarity:     94.31%        Conservative:   38
Best Local Similarity:  89.04%        Mismatches:     33
Query Match:            74.77%        Indels:         8
DB:                     1             Gaps:           4
```

US-09-610-313B-32 (1-2457) x GNVWA2 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------GACAACCCC 333
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db         61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProVal 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db        361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMet 400

Qy       1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db        461 AsnArgGluIleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeu 480

Qy       1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db        481 ValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500
```

```
Qy     1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db      501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy     1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db      521 LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys 540

Qy     1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db      541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyr 560

Qy     1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db      561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy     1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db      581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy     1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db      601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620

Qy     2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db      621 ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy     2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
Db      641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy     2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
Db      661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy     2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
Db      681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy     2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
Db      701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720

Qy     2356 GAT 2358
Db      721 Asp 721
```

RESULT 8
GNVWVL
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LV)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C;Accession: A03967
R;Muesing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.
Nature 313, 450-458, 1985
A;Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi
A;Reference number: A93355; MUID:85111157; PMID:2982104
A;Accession: A03967
A;Molecule type: DNA
A;Residues: 1-1012 <MUE>
A;Cross-references: UNIPROT:P03368
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle
F;66-164/Product: retropepsin #status predicted <RTP>
F;90/Active site: Asp (shared with dimeric partner) #status experimental

Alignment Scores:
Pred. No.:   4.53e-151   Length: 1012

```
Score:                    3396.00     Matches:        643
Percent Similarity:       93.12%      Conservative:   34
Best Local Similarity:    88.45%      Mismatches:     30
Query Match:              74.56%      Indels:         20
DB:                       1           Gaps:           5

US-09-610-313B-32 (1-2457) x GNVWVL (1-1012)

Qy      238 GCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------------------------ 267
Db        4 AlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerPro 23

Qy      268 ------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG--- 318
Db       24 ThrIleSerSerGluGlnThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrp 43

Qy      319 ---CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------ 369
Db       44 GlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPhe 63

Qy      370 AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAG 429
Db       64 AsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGln 83

Qy      430 ATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTG 489
Db       84 LeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeu 103

Qy      490 CCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAG 549
Db      104 ProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGln 123

Qy      550 TACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGC 609
Db      124 TyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIleGlyThrValLeuValGly 143

Qy      610 CCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAAC 669
Db      144 ProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsn 163

Qy      670 TTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCC 729
Db      164 PheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyPro 183

Qy      730 AAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAG 789
Db      184 LysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThr 203

Qy      790 GAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCC 849
Db      204 GluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrPro 223

Qy      850 GTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAG 909
Db      224 ValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGlu 243

Qy      910 CTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGC 969
Db      244 LeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGly 263

Qy      970 CTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCC 1029
Db      264 LeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValPro 283

Qy     1030 CTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACC 1089
Db      284 LeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThr 303

Qy     1090 CCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATC 1149
Db      304 ProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIle 323

Qy     1150 TTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTG 1209
```

```
                 ||||||||||||||||||||||||||||||||||||||||    :::|||||:::||||||
Db         324 PheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleVal 343

Qy        1210 ATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGC 1263
               |||||||||      ||||||||||||||||||||||||||||||||||||||||||||||
Db         344 IleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArg 363

Qy        1264 GCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAG 1323
                  ||||||||||||||||:::||||||||||||||||||      |||||||||||||||
Db         364 ThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLys 383

Qy        1324 AAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGG 1377
               |||||||||||||||||||||||||||            ||||||||||||||||||||||
Db         384 LysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrp 403

Qy        1378 ACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAG 1437
               ||||||||||||||||    ||||||||||||:::||||||||||||||||||||||||||
Db         404 ThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLys 423

Qy        1438 CTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTG 1497
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         424 LeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeu 443

Qy        1498 TGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCC 1557
               ||||||||||||||||||     |||||||||||||::::::::||||||||||||||||||
Db         444 CysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAla 463

Qy        1558 GAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTAC 1617
               |||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db         464 GluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyr 483

Qy        1618 GACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTAC 1677
               ||||||||||||||||||:::||||||||||||||||||||||||       ||||||||||||
Db         484 AspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyr 503

Qy        1678 CAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACC 1737
               ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db         504 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 523

Qy        1738 GCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGC 1797
               ||||||||||||||||||||||||||||||||||||||||||||||||||      ||||||
Db         524 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSer 543

Qy        1798 ATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAG 1857
               ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db         544 IleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGluThrTrpGlu 563

Qy        1858 ACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACC 1917
               ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db         564 ThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThr 583

Qy        1918 CCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACC 1977
               ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db         584 ProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThr 603

Qy        1978 TTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACC 2037
               ||||||||||||||||||||||||||||||||||||:::::::||||||||||||||||||||
Db         604 PheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGlyLysAlaGlyTyrValThr 623

Qy        2038 GACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTG 2097
               ::::::|||||||||||||:::|||    ||||||    |||||||||||||||||||||||||
Db         624 AsnLysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeu 643

Qy        2098 CAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGC 2157
               |||||||||    ||||||||||||||||||||||||    ||||||||||||||||||||||||
Db         644 GlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSer 663

Qy        2158 CAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAAC 2217
               ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db         664 GlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsn 683

Qy        2218 CAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCAC 2277
               ||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db         684 GlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHis 703

Qy        2278 AAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2337
               |||||||||||||||||||||||:::|||||||||||||||    ||||||||||||:::
Db         704 LysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysIle 723

Qy        2338 CTGTTCCTGGACGGCATCGAT 2358
               |||||||||||||||||||||
Db         724 LeuPheLeuAspGlyIleAsp 730
```

RESULT 9
GNLJSI
HIV-1 retropepsin (EC 3.4.23.16) - simian immunodeficiency virus SIVcpz
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: simian immunodeficiency virus SIVcpz
A;Note: host Pan troglodytes (chimpanzee)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: S09984
R;Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.
Nature 345, 356-359, 1990
A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
A;Reference number: S09983; MUID:90259077; PMID:2188136
A;Accession: S09984
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1027 <HUB>
A;Cross-references: UNIPROT:P17283; EMBL:X52154
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;81-180/Product: retropepsin #status predicted <RTP>
F;105/Active site: Asp (shared with dimeric partner) #status predicted

```
Alignment Scores:
Pred. No.:              1.93e-140     Length:         1027
Score:                  3168.00       Matches:        591
Percent Similarity:     88.38%        Conservative:   63
Best Local Similarity:  79.86%        Mismatches:     79
Query Match:            69.55%        Indels:         7
DB:                     1             Gaps:           3

US-09-610-313B-32 (1-2457) x GNLJSI (1-1027)

Qy         159 CTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAA 218
               ||||||   |||           |||   |||:::|||  |||||||||       |||:::
Db           7 LeuLeuAlaValTrpAlaArgGlyThrProAsnGluArgLeuHisArgLysThrGlyGlu 26

Qy         219 CTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCA 278
                |||||||||||      ||||||||||||||   :::||||||:::      :::|||||
Db          27 -PhePheArgGluArgLeuAlaPheProGlnArgGluAlaArgGlnLeuCysAlaGluGl 46

Qy         279 GAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------GACAACCC 332
               ||||||   |||   ||||||   |||||||||   |||   |||       :::   ||
Db          46 nAsnArgThrAsnGlyProThrAspArgGluLeuTrpValProGlyGlyArgGluGluPr 66

Qy         333 CCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTG 392
               |      |||   |||   |||:::       :::   |||   ||||||||||||||||
Db          66 oGlyGluGluArgGlyArgGluGlnSerIleSerThrAsnLeuProGlnIleThrLeuTr 86

Qy         393 GCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACAC 452
               ||||||||||||:::    :::||||||   ||||||:::   ||||||||||||||||
Db          86 pGlnArgProLeuIleProValLysValGluGlyGlnLeuCysGluAlaLeuLeuAspTh 106

Qy         453 CGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGAT 512
               ||||||||||||||||||:::|||   :::   |||   |||   |||||||||||||||
Db         106 rGlyAlaAspAspThrValIleGluArgIleGlnLeuGlnGlyLeuTrpLysProLysMe 126

Qy         513 GATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGAT 572
               |||||||||||||||||||||||||||||||:::|||:::|||      :::   ||||||||
Db         126 tIleGlyGlyIleGlyGlyPheIleLysValLysGlnPheAspAsnValHisIleGluIl 146
```

```
Qy        573 CTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGG 632
              |   |||:::|||   :::|||||||||||:::|||||||||||||||||||||||||
Db        146 eGluGlyArgLysValValGlyThrValLeuValGlyProThrProValAsnIleIleGl 166

Qy        633 CCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGAC 692
              ||||||:::||||||||||||||||||||||||    |||||||||||   ||||||||
Db        166 yArgAsnIleLeuThrGlnLeuGlyCysThrLeuValPheProIleSerSerIleGluTh 186

Qy        693 CGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGAC 752
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:: 
Db        186 rValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuSe 206

Qy        753 CGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGAT 812
              :   ||||||||||||||||||    ||||||:::||||||||||||||||||||
Db        206 rAlaGluLysIleLysAlaLeuThrGluIleCysGlnGluMetGluLysGluGlyLysIl 226

Qy        813 CACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGA 872
              |:::|||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        226 eSerLysIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAs 246

Qy        873 CAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTT 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 pSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPh 266

Qy        933 CTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGAC 992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 eTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValTh 286

Qy        993 CGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTA 1052
              |||||||||||||||||||||||||||||    ||||||||:::|||||||||||||
Db        286 rValLeuAspValGlyAspAlaTyrPheSerCysProLeuAspLysAspPheArgLysTy 306

Qy       1053 CACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAA 1112
              |||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db        306 rThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyValArgTyrGlnTyrAs 326

Qy       1113 CGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGAT 1172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 nValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSerMetThrLysIl 346

Qy       1173 CCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCT 1226
              ||||||||||||||||    :::||||||:::|||   |||||||||         ||
Db        346 eLeuGluProPheArgGluLysAsnProAspIleThrIleTyrGlnTyrMetAspAspLe 366

Qy       1227 GTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAA 1286
              ||||||||||||||||||||||||   |||||||||    |||:::|||||||||||||:: 
Db        366 uTyrValGlySerAspLeuGluIleAspGlnHisArgLysLysValGluGluLeuArgGl 386

Qy       1287 GCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTT 1346
              :||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 nHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPh 406

Qy       1347 CCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCC 1400
              ||||            ||||||||||||||||||||||||||||||||||:::||||||
Db        406 eLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuPr 426

Qy       1401 CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGC 1460
              ||||||||||    |||||||||||||||||||||||||:::||||||||||||||||||
Db        426 oGluLysGluValTrpThrValAsnAspIleGlnLysLeuIleGlyLysLeuAsnTrpAl 446

Qy       1461 CAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAA 1520
              ||||||||||||||||||||||||::::::||||||||||||||||:::||||||    ||
Db        446 aSerGlnIleTyrProGlyIleLysIleLysGlnLeuCysLysLeuIleArgGlyThrLy 466

Qy       1521 GGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCG 1580
              |    ||||||||:::||||||||||||    ||||||||||||||||||||||||||||
Db        466 sLysLeuThrAspValValProLeuThrProGluAlaGluLeuGluLeuAlaGluAsnAr 486

Qy       1581 CGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGC 1640
              |||||||:::      |||||||||||||||||||||||||    |||:::|||:::||
Db        486 gGluIleValSerThrProValHisGlyValTyrTyrAspProAspLysGluLeuIleAl 506

Qy       1641 CGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAA 1700
              ||||||||||||||||||:::   |||||||||||||||||:::|||||||||    ||
Db        506 aGluIleGlnLysGlnGlyAsnCysGlnTrpThrTyrGlnIlePheGlnGluProHisLy 526

Qy       1701 GAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCA 1760
              ||||||||||||||||||||||||:::    |||:::|||||||||||||||::::::||
Db        526 sAsnLeuLysThrGlyLysTyrAlaArgGlnArgSerAlaHisThrAsnAspIleArgGl 546

Qy       1761 GCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCC 1820
              ||||   ||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        546 nLeuAlaGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThrPr 566

Qy       1821 CAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCA 1880
              |||||||||||||||:::||||||||:::||||||   ||||||    :::||||||||
Db        566 oLysPheArgLeuProValGlnLysGluSerTrpGluAlaTrpTrpAlaGluTyrTrpGl 586

Qy       1881 GGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTA 1940
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        586 nAlaThrTrpIleProGluTrpGluPheIleAsnThrProProLeuValLysLeuTrpTy 606

Qy       1941 CCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAA 2000
              |    ||||||    |||||||||           :::|||:::||||||||||||||||||
Db        606 rSerLeuGluThrGluProIleProThrThrAspThrTyrTyrValAspGlyAlaAlaAs 626

Qy       2001 CCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGT 2060
              ||||||||||||    |||||||||||||||||||||:::|||:::||||||||||::
Db        626 nArgGluThrLysThrGlyLysAlaGlyTyrValThrAspLysGlyLysGlnLysIleIl 646

Qy       2061 GAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCA 2120
              :||||||         ||||||||||||:::     ||||||:::|||:::     ||||||||||
Db        646 eSerLeuGluAsnThrThrAsnGlnGlnAlaGluLeuLysAlaLeuLeuLeuAlaLeuGl 666

Qy       2121 GGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCA 2180
              ||||||      :::||||||||||||||||||||||||   ||||||||||||||
Db        666 nAspSerAspGlnGlnValAsnIleValThrAspSerGlnTyrValLeuGlyIleIleGl 686

Qy       2181 GGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAA 2240
              |:::||||||||    ||||||||||||||||||||||||||||:::|||||||
Db        686 nSerGlnProAspHisSerGluSerGluLeuValAsnGlnIleIleGluGluLeuIleLy 706

Qy       2241 GAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCA 2300
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db        706 sLysGluLysIleTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGl 726

Qy       2301 GATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |:::|||||||||||||    |||||||||||||||||||||||||||||||||||||
Db        726 nValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 745
```

RESULT 10
S33980
pol polyprotein - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C;Accession: S33980
R;Carlini, F.
submitted to the EMBL Data Library, November 1991
A;Reference number: S33979
A;Accession: S33980
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-912 <CAR>
A;Cross-references: UNIPROT:Q78241; EMBL:Z11530; NID:g60192; PIDN:CAA77622.1; PID:g60194
C;Superfamily: pol polyprotein

```
Alignment Scores:
Pred. No.:                1.33e-135        Length:          912
Score:                    3064.00          Matches:         572
Percent Similarity:       95.40%           Conservative:    29
Best Local Similarity:    90.79%           Mismatches:      25
Query Match:              67.27%           Indels:          4
DB:                       2                Gaps:            2
```

US-09-610-313B-32 (1-2457) x S33980 (1-912)

```
Qy    481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db      1 MetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 20

Qy    541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db     21 ValLysGlnTyrAspGlnIleAlaIleGluIleCysGlyHisLysAlaIleGlyThrVal 40

Qy    601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db     41 LeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCys 60

Qy    661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db     61 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 80

Qy    721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db     81 AspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuIleGlu 100

Qy    781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db    101 IleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyr 120

Qy    841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db    121 AsnThrProValPheAlaIleLysLysLysAspGlyThrLysTrpArgLysLeuValAsp 140

Qy    901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db    141 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 160

Qy    961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db    161 ProGlyGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 180

Qy   1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db    181 SerValProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsn 200

Qy   1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db    201 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 220

Qy   1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db    221 ProAlaIlePheGlnAlaSerMetThrLysIleLeuGluProPheArgLysGlnAsnPro 240

Qy   1201 GAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
Db    241 GluIleIleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 260

Qy   1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Db    261 GlnHisArgThrLysIleGluGluLeuArgArgHisLeuLeuArgTrpGlyPheThrThr 280

Qy   1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATC------GAGCTGCACCCC 1368
Db    281 ProAspLysLysHisGlnLysGluProProPheLeuTrpIleGlyTyrGluLeuHisPro 300

Qy   1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
Db    301 AspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAsp 320

Qy   1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
Db    321 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLysVal 340

Qy   1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
Db    341 ArgGlnLeuCysArgLeuLeuArgGlyAlaLysAlaLeuThrGluValIleProLeuThr 360

Qy   1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
Db    361 LysGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysThrProValHisGly 380

Qy   1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
Db    381 ValTyrTyrAspProSerLysAspLeuValAlaGluIleGlnLysGlnGlyLeuGlyGln 400

Qy   1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
Db    401 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 420

Qy   1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
Db    421 MetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 440

Qy   1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
Db    441 ThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGlu 460

Qy   1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
Db    461 ThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 480

Qy   1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
Db    481 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 500

Qy   1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
Db    501 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly 520

Qy   2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
Db    521 TyrValThrAspArgGlyArgGlnLysValIleSerLeuThrAspThrThrAsnGlnLys 540

Qy   2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
Db    541 ThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleVal 560

Qy   2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
Db    561 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 580

Qy   2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
Db    581 LeuValSerGlnIleIleGluHisLeuIleLysLysGluLysValTyrLeuAlaTrpVal 600

Qy   2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
Db    601 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIle 620

Qy   2329 CGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db    621 ArgLysValLeuPheLeuAspGlyIleAsp 630
```

RESULT 11
T01668
pol polyprotein - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T01668
R;Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.
Cell 46, 63-74, 1986
A;Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol
A;Reference number: Z14389; MUID:86245056; PMID:2424612
A;Accession: T01668
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-902 <ALI>
A;Cross-references: EMBL:K03456; NID:g60228; PIDN:CAA28012.1; PID:g60230
C;Superfamily: pol polyprotein

Alignment Scores:

```
Pred. No.:              5.37e-135     Length:         902
Score:                  3051.00       Matches:        570
Percent Similarity:     96.45%        Conservative:   28
Best Local Similarity:  91.94%        Mismatches:     18
Query Match:            66.98%        Indels:         4
DB:                     2             Gaps:           2

US-09-610-313B-32 (1-2457) x T01668 (1-902)

Qy        511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 20

Qy        571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
              |||||||||||||||||||||||||||:::|||:::||||||||||||||||||||||||
Db         21 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 40

Qy        631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         41 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 60

Qy        691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db         61 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 80

Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
              |||||||||||||||||||||||||||   ||||||::::::||||||||||||||||||
Db         81 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 100

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 120

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        121 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 140

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 160

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 180

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 200

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        201 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 220

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
              ||||||||||||||||||   :::||||||||||||||||||||||||            
Db        221 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 240

Qy       1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        241 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 260

Qy       1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
              :::|||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 280

Qy       1345 TTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1398
              ||||||            ||||||||||||||||||||||||||||||||||||:::|||
Db        281 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 300

Qy       1399 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1458
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 320

Qy       1459 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1518
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        321 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 340

Qy       1519 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1578
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        341 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 360

Qy       1579 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1638
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||:::
Db        361 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 380

Qy       1639 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1698
              |||||||||||||||||||||      ||||||||||||||||||||||||      :::
Db        381 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 400

Qy       1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1758
              ||||||||||||||||||||||||::::::::::::||||||||||||||||||||||||
Db        401 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 420

Qy       1759 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1818
              ||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db        421 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 440

Qy       1819 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878
              |||||||||||||||||||||||||||||||||   ||||||||||||||||:::||||||
Db        441 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 460

Qy       1879 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1938
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 480

Qy       1939 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 1998
              ||||||||||||   |||||||||:::|||||||||||||||||||||||||||||||||
Db        481 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 500

Qy       1999 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2058
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||:::
Db        501 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 520

Qy       2059 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        521 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 540

Qy       2119 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 560

Qy       2179 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2238
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        561 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 580

Qy       2239 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2298
              :::|||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 600

Qy       2299 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||:::|||||||||         ||||||||||||||||||||||||||||||||||||
Db        601 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 620
```

RESULT 12
B47175
reverse transcriptase, AZT-sensitive variant - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: B47175
R;Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993
A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in th
A;Reference number: A47175; MUID:93126353; PMID:7678340
A;Accession: B47175
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: DNA

```
A;Residues: 1-559 <MOH>
A;Cross-references: UNIPROT:Q9PXX1
A;Note: sequence extracted from NCBI backbone (NCBIP:122099)
C;Superfamily: pol polyprotein

Alignment Scores:
Pred. No.:                5.7e-120       Length:         559
Score:                    2728.00        Matches:        510
Percent Similarity:       95.35%         Conservative:   23
Best Local Similarity:    91.23%         Mismatches:     22
Query Match:              59.89%         Indels:         4
DB:                       2              Gaps:           2

US-09-610-313B-32 (1-2457) x B47175 (1-559)

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 20

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              |||||||||||||||||||||||||||||||||||||||||||    ||||||    |||
Db         21 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGlu 40

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db         41 MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal 60

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 80

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 100

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 120

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 140

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        141 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 160

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||    :::||||||:::||||||||||
Db        161 GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleValIle 180

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||            ||||||||||||||||||||||||||||||||||||||
Db        181 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr 200

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||:::|||||||||||||||||||    ||||||||||||||||
Db        201 LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLys 220

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||                ||||||||||||||||||||
Db        221 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 240

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||    |||||||||||||:::||||||||||||||||||||||||||||
Db        241 ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 260

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 280

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              |||||||||||||||   ||||||||||||:::::::::|||||||||||||||||||||
Db        281 LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGlu 300

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        301 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 320

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||:::|||||||||||||||||||||||      ||||||||||||||
Db        321 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGln 340

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||   |||
Db        341 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla 360

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        361 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle 380

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              |||||||||||||||||||:::|||:::||||||||||||||||||||||||||||||||
Db        381 ValIleTrpGlyLysThrProArgPheLysLeuProIleGlnLysGluThrTrpGluThr 400

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 TrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 420

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        421 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 440

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||:::
Db        441 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 460

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              :::|||||||||:::|||   ||||||   ||||||||||||||||||||||||||||||
Db        461 LysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeuGln 480

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||    ||||||||||||||||||||||   |||||||||||||||||||||||||
Db        481 AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln 500

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        501 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln 520

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        521 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys 540

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2337
              ||||||||||||||||||||:::||||||||||||||||   |||||||||||||||
Db        541 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 559

RESULT 13
A47175
reverse transcriptase, AZT-resistant variant - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A47175
R;Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993
A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the
A;Reference number: A47175; MUID:93126353; PMID:7678340
A;Accession: A47175
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: nucleic acid
A;Residues: 1-559 <MOH>
A;Cross-references: UNIPROT:Q9PXX2
A;Note: sequence extracted from NCBI backbone (NCBIP:122100)
C;Superfamily: pol polyprotein

Alignment Scores:
```

```
Pred. No.:                  4.88e-119      Length:        559
Score:                      2708.00        Matches:       506
Percent Similarity:         95.17%         Conservative:  26
Best Local Similarity:      90.52%         Mismatches:    23
Query Match:                59.45%         Indels:        4
DB:                         2              Gaps:          2

US-09-610-313B-32 (1-2457) x A47175 (1-559)

Qy    673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
Db      1 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 20

Qy    733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
Db     21 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGlu 40

Qy    793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db     41 MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal 60

Qy    853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db     61 PheAlaIleLysLysLysAspSerThrArgTrpArgLysLeuValAspPheArgGluLeu 80

Qy    913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db     81 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 100

Qy    973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db    101 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 120

Qy   1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db    121 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 140

Qy   1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db    141 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 160

Qy   1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db    161 GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspMetValIle 180

Qy   1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db    181 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr 200

Qy   1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db    201 LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 220

Qy   1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
Db    221 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 240

Qy   1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
Db    241 ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 260

Qy   1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
Db    261 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys 280

Qy   1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
Db    281 LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleGlnLeuThrGluGluAlaGlu 300

Qy   1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
Db    301 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 320

Qy   1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
Db    321 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGln 340

Qy   1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
Db    341 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla 360

Qy   1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
Db    361 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle 380

Qy   1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
Db    381 ValIleTrpGlyLysIleProArgPheLysLeuProIleGlnLysGluThrTrpGluAla 400

Qy   1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
Db    401 TrpTrpIleGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 420

Qy   1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
Db    421 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 440

Qy   1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
Db    441 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 460

Qy   2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
Db    461 LysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGln 480

Qy   2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
Db    481 AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln 500

Qy   2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
Db    501 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln 520

Qy   2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
Db    521 IleIleGluGluLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys 540

Qy   2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTG 2337
Db    541 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 559

RESULT 14
S46347
pol polyprotein - simian immunodeficiency virus SIVagm (isolate SAB-1)
C;Species: simian immunodeficiency virus SIVagm
A;Variety: isolate SAB-1
C;Date: 25-Dec-1994 #sequence_revision 14-Feb-1997 #text_change 26-Aug-1999
C;Accession: S46347
R;Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.; Barre-Sinoussi, F.; Hirsch, V.M.;
EMBO J. 13, 2935-2947, 1994
A;Title: Mosaic genome structure of simian immunodeficiency virus from West African grec
A;Reference number: S46335; MUID:94298785; PMID:8026477
A;Accession: S46347
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1039 <JIN>
A;Cross-references: EMBL:U04005; NID:g466229; PIDN:AAA21505.1; PID:g466231
A;Experimental source: isolate SAB-1; sabaeus monkey
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
A;Note: this reading frame extends between two stop codons and does not begin with a sta
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein

Alignment Scores:
Pred. No.:             2.61e-100      Length:        1039
```

```
Score:                    2307.00      Matches:        436
Percent Similarity:       73.54%       Conservative:   120
Best Local Similarity:    57.67%       Mismatches:     154
Query Match:              50.65%       Indels:         46
DB:                       2            Gaps:           8

US-09-610-313B-32 (1-2457) x S46347 (1-1039)

Qy    220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC------AAGGCCCGCGAGTTCCCCAGC 273
Db      1 PhePheArg---------ValTrpProLeuGlyGlnArgGluThrGlnGluPheProSer 17

Qy    274 GAGCAGAACCGCGCCAAC---AGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGC------ 324
Db     18 AspLeuHisGlnThrAsnSerSerProAsnGlyThrGlyLeuGlnGlnAlaGlyGlyLys 37

Qy    325 ------------------GACAACCCCCGCAGCGAGGCC--------------------- 345
Db     38 LeuValCysArgGlnThrSerAspGlnArgThrArgAlaArgArgSerSerAsnSerPro 57

Qy    346 ---------------------------GGCGCCGAGCGC--------------------- 357
Db     58 ValLysAlaValCysCysSerGlyGluThrAlaGluThrAlaValAlaLysProLeuAla 77

Qy    358 ---------------CAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
Db     78 ThrThrGluProLeuArgGlyGlyLeuGlnLeuProGlnValSerLeuTrpArgArgPro 97

Qy    403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
Db     98 MetLysThrValTyrIleGluGlyGlnLysValThrAlaLeuLeuAspThrGlyAlaAsp 117

Qy    463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
Db    118 AspSerValIleGlnGlyIleGluLeuGlyAspAsnTrpLysProArgIleIleGlyGly 137

Qy    523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
Db    138 IleGlyGlyCysIleAsnValLysAlaTyrHisAsnGlnGluValLysIleGluAspLys 157

Qy    583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
Db    158 ThrCysLysAlaThrIleLeuValGlyGluThrProValAsnIleIleGlyArgAsnVal 177

Qy    643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
Db    178 LeuAlaGlnLeuGlyValThrLeuAsnLeuThrGlnArgGluIleGluProIleLysVal 197

Qy    703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
Db    198 HisLeuLysProGlyGlnAspGlyProArgIleArgGlnTrpProLeuSerLysGluLys 217

Qy    763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
Db    218 IleGluAlaLeuLysAlaIleCysGluAspLeuGluLysGlnGlyHisLeuGluArgIle 237

Qy    823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
Db    238 GlyProGluAsnProTyrAsnThrProValPheAlaIleArgLysLysAspLysThrGln 257

Qy    883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
Db    258 TrpArgIleLeuMetAspPheArgGlnLeuAsnLysSerThrGlnAspPheGlnGluVal 277

Qy    943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
Db    278 GlnLeuGlyIleProHisProAlaGlyLeuGlnGlnArgGluGlnIleThrValLeuAsp 297

Qy   1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
Db    298 IleGlyAspAlaTyrPheSerCysProLeuAspProAspPheGlnLysTyrThrAlaPhe 317

Qy   1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
Db    318 ThrIleProSerValAsnAsnArgGluProGlyIleArgTyrGlnTyrLysValLeuPro 337

Qy   1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
Db    338 GlnGlyTrpLysGlySerProThrIlePheGlnThrThrAlaAsnLysIleLeuGlnGlu 357

Qy   1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGC 1236
Db    358 PheArgGlnLysAsnProAspValAspIleTyrGlnTyrMetAspAspMetLeuIleAla 377

Qy   1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
Db    378 SerAspArgProLysAlaGluHisLeuValMetValGlnGlnLeuArgAspTyrLeuGlu 397

Qy   1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------ 1350
Db    398 ThrTrpGlyPheLysThrProGluLysLysPheGlnLysAspProProTyrLeuTrpMet 417

Qy   1351 CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1410
Db    418 GlyTyrGluLeuTyrProLysLysTrpGlnLeuGlnGluIleThrLeuProGluArgGlu 437

Qy   1411 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1470
Db    438 GluTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 457

Qy   1471 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1530
Db    458 TyrThrGlyIleLysThrLysHisLeuCysArgLeuIleArgGlyAlaArgProLeuThr 477

Qy   1531 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1590
Db    478 GluIleValGlnTrpThrGluGluAlaGluLeuGluLeuGluGluAsnArgGlnIleLeu 497

Qy   1591 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1650
Db    498 ArgGlnLysGlnGlnGlyGlnTyrTyrAspProAlaLeuProLeuArgAlaLysValLeu 517

Qy   1651 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1710
Db    518 LysLeuGlyAspGlyGlnTrpGlyTyrGlnIleTyrGlnProGluAsnLysIleLeuLys 537

Qy   1711 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1770
Db    538 ValGlyLysTyrAlaLysIleLysThrAlaHisThrAsnGluLeuArgMetLeuAlaGly 557

Qy   1771 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1830
Db    558 LeuValGlnLysIleGlyLysGluSerIleValIleTrpGlyGlnIleProIleMetGlu 577

Qy   1831 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1890
Db    578 LeuProValGluArgGluLeuTrpGluGlnTrpTrpSerAspTyrTrpGlnValThrTrp 597

Qy   1891 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1950
Db    598 IleProGluTrpGluMetValSerThrProGlnLeuIleArgLeuTrpTyrLysLeuVal 617

Qy   1951 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2010
Db    618 LysAspProIleProGlyGluAlaValTyrTyrValAspGlyAlaAlaAsnArgAsnSer 637

Qy   2011 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2070
Db    638 LysGluGlyLysAlaGlyTyrLeuThrAspArgGlyAspGlnLysValValAlaLeuGlu 657

Qy   2071 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2130
Db    658 AsnThrThrAsnGlnLysAlaGluLeuGluAlaIleLeuLeuAlaLeuArgAspSerGly 677

Qy   2131 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2190
```

```
Db        678 SerLysValAsnIleIleThrAspSerGlnTyrAlaMetGlyIleIleAlaGlyGluPro 697

Qy       2191 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2250
                 :::|||::::::    :::|||   |||||||||||:::||||||||||||||
Db        698 ThrGluSerAspAsnAsnIleValGlnGlnIleIleGluGluLeuIleLysLysGluAla 717

Qy       2251 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2310
              ||||||:::::|||||||||||||||||||:::||||||||||||:::||||||||
Db        718 ValTyrIleAlaTrpValProAlaHisLysGlyValGlyGlyAsnGluGluIleAspLys 737

Qy       2311 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||||||||:::|||||||||:::||||||||||||||    |||:::
Db        738 LeuValSerGlnGlyIleArgGlnValLeuPheLeuAspArgIleGlu 753
```

RESULT 15
S53092
pol polyprotein - human immunodeficiency virus type 2
C;Species: human immunodeficiency virus type 2, HIV-2
C;Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C;Accession: S53092
R;Becker, M.; Zorr, B.; Becker, A.; Habermehl, K.O.
submitted to the EMBL Data Library, March 1995
A;Description: Molecular and phylogenetic characterisation of a Guinea-Bissau-derived hu
A;Reference number: S53091
A;Accession: S53092
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1055 <BEC>
A;Cross-references: UNIPROT:O73194; EMBL:Z48731
C;Superfamily: pol polyprotein
C;Keywords: polyprotein

Alignment Scores:

Pred. No.:	6.37e-94	Length:	1055
Score:	2170.00	Matches:	413
Percent Similarity:	70.21%	Conservative:	129
Best Local Similarity:	53.50%	Mismatches:	187
Query Match:	47.64%	Indels:	43
DB:	2	Gaps:	11

US-09-610-313B-32 (1-2457) x S53092 (1-1055)

```
Qy        156 GGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCAC---CGAGCGCCA 212
              ||||||||||:::   |||             |||        |||   :::   |||
Db          3 GlyLeuLeuGluMetTrpGln------------AspArgThrTyrHisGlyLysValPro 18

Qy        213 GGCCAACTT------------------------------------------CTTCCG 227
                 |||                                                ||||||
Db         19 ArgGlnThrGlyGlyPhePheArgAspTrpProLeuGlyLysGluAlaProGlnLeuPro 38

Qy        228 CGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAA------ 281
              ||||||||||||                               |||   |||:::
Db         39 ArgGlyProGlySerAlaGlyAlaAsnThrAsnSerThrProSerArgSerSerSerGly 58

Qy        282 CCGCGCCAACAGCCCCACCAGCCGCGA------GCTGCAGGTGCGCGGCGACAACCCCCG 335
              |||      :::   :::       |||      |||   ||||||   |||:::
Db         59 ProThrGlyGluIleTyrAlaAlaArgGluLysAlaGluGlyAlaGluArgGluThrIle 78

Qy        336 CAGCGAGGCCGGCGC------CGAGCGCCA------GGGCACCCTGAACTT--------- 374
              |||||||||   |||            ||||||     |||            :::
Db         79 GlnArgGlyAspArgGlyLeuAlaAlaProArgAlaGlyLysAspThrMetGlnGlyAsp 98

Qy        375 ------------------CCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAA 416
                                 ||| |||   :::||||||    ||||||:::|||:::
Db         99 AsnArgGlyPheAlaAlaPro-GlnPheSerLeuTrpAsnArgProValValThrAlaHi 118

Qy        417 GGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGA 476
               :::   ||||||     |||   ||||||||||||||||||||||||:::::::::
Db        118 sIleGluGlyGlnProValGluValLeuLeuAspThrGlyAlaAspAspSerIleValAl 138

Qy        477 GGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
                :::    |||           :::    ||||||::::::||||||||||||||||||
Db        138 aGlyIleGluLeuGlySerAsnTyrSerProLysIleValGlyGlyIleGlyGlyPheIl 158

Qy        537 CAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAC 596
              |      ::::::|||      :::    ||||||:::    ||||||:::        ||
Db        158 eAsnThrLysGluTyrLysAsnValGluIleGluValLeuGlyLysArgValArgAlaTh 178

Qy        597 CGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGG 656
              |::::::   |||   ||||||:::||||||    |||||||||:::||||||    |||||
Db        178 rIleMetThrGlyAspThrProIleAsnIlePheGlyArgAsnIleLeuThrAlaLeuGl 198

Qy        657 CTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGG 716
              |    :::||||||   |||::::::   ||||||    :::    :::    |||||||||
Db        198 yMetSerLeuAsnLeuProValAlaLysIleGluProIleLysIleMetLeuLysProGl 218

Qy        717 CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776
              |    |||||||||||||::::::|||||||||||||||:::|||||||||:::|||||
Db        218 yLysAspGlyProLysLeuArgGlnTrpProLeuThrLysGluLysIleGluAlaLeuLy 238

Qy        777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
                 |||||||||:::||||||:::||||||:::::    :::         |||   |||||
Db        238 sGluIleCysGluLysMetGluArgGluGlyGlnLeuGluGluAlaProProThrAsnPr 258

Qy        837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
                 ||||||||||||   ||||||||||:::||||||     ||||||||||   |||::
Db        258 oTyrAsnThrProThrPheAlaIleArgLysLysAspLysAsnLysTrpArgMetLeuIl 278

Qy        897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
              :||||||||||||||||||   ||||||||||||   |||:::|||||||||||||
Db        278 eAspPheArgGluLeuAsnLysValThrGlnAspPheThrGluIleGlnLeuGlyIlePr 298

Qy        957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
              |||||||||||||||   ||||||:::   :::||||||||||||||||||||||||||||
Db        298 oHisProAlaGlyLeuAlaLysLysArgArgIleThrValLeuAspValGlyAspAlaTy 318

Qy       1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
              ||||||:::||||||   |||||||||||:::|||||||||||||||||:::||||||::
Db        318 rPheSerIleProLeuHisGluAspPheArgGlnTyrThrAlaPheThrLeuProSerVa 338

Qy       1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
              :||||||     ||||||   ||||||   |||   ||||||||||||||||||||||||
Db        338 lAsnAsnAlaGluProGlyLysArgTyrIleTyrLysValLeuProGlnGlyTrpLysGl 358

Qy       1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
              ||||||:::||||||||     :::|||    ::::::|||||||||||||||      ||
Db        358 ySerProAlaIlePheGlnTyrThrMetArgGlnValLeuGluProPheArgLysAlaAs 378

Qy       1197 CCCCGAGATCGTGATC------TACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250
              |    ::::::::|||      |||             :::    :::    ||||||
Db        378 nGlnAspValIleIleIleGlnTyrMetAspAspIleLeuIleAlaSerAspArgThrAs 398

Qy       1251 CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAC 1310
                :::|||         :::    :::|||::::::    |||          ||||||::
Db        398 pLeuGluHisAspArgValValLeuGlnLeuLysGluLeuLeuAsnSerLeuGlyPheSe 418

Qy       1311 CACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTC------CTGCCCATCGAGCTGCA 1364
              :||||||||::::|||  ||||||:::||||||:::          :::      ||||||
Db        418 rThrProAspGluLysPheGlnLysAspProProTyrArgTrpMetGlyTyrGluLeuTr 438

Qy       1365 CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAA 1424
                |||   ||||||   :::|||   |||:::||||||:::||||||   ||||||||||||
Db        438 pProThrLysTrpLysLeuGlnLysIleGlnLeuProGlnLysGluValTrpThrValAs 458

Qy       1425 CGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAA 1484
              ||||||||||||||||||||   |||||||||||:::||||||||||||||||||||
Db        458 nAspIleGlnLysLeuValGlyValLeuAsnTrpAlaAlaGlnIleTyrProGlyIleLy 478

Qy       1485 GGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCT 1544
              |    :::    ||||||:::|||:::||||||               ||||||:::   :::
Db        478 sThrLysHisLeuCysArgLeuIleArgGlyLysMetThrLeuThrGluGluIleGlnTr 498

Qy       1545 GACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA 1604
                ||||||    ||||||   ||||||   |||||||||   ||||||         :::
Db        498 pThrGluLeuAlaGluAlaGluLeuGluGluAsnArgValIleLeuSerGlnGluGlnGl 518
```

```
Qy       1605 CGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGA 1664
               |||  ||||||      |||:::|||  |||   :::||||||          ::
Db        518 uGlyHisTyrTyrGlnGluGluLysGluLeuGluAlaThrValGlnLysAspGlnAspAs 538

Qy       1665 CCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 1724
              :|||||||||||:::::::::|||         |||  ||||||  ||||||||||
Db        538 nGlnTrpThrTyrLysValHisGlnGlyGlu---LysThrLeuLysValGlyLysTyrAl 557

Qy       1725 CAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGAT 1784
              ||||::::::        ||||||||  |||:::  |||   :::  |||||||||
Db        557 aLysValLysAsnThrHisThrAsnGlyValArgLeuLeuAlaGlnValValGlnLysIl 577

Qy       1785 CGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAA 1844
              |        |||::::::|||||||||||:::   |||||||||   ||||||::::::::
Db        577 eGlyLysGluAlaLeuValIleTrpGlyArgIleProLysPheHisLeuProValGluAr 597

Qy       1845 GGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGA 1904
              :|||  ||||||  ||||||  |||||||||||  ||||||||||||:::|||::
Db        597 gGluIleTrpGluGlnTrpTrpAspAspTyrTrpGlnValThrTrpIleProAspTrpAs 617

Qy       1905 GTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT 1964
              :||||||:::|||||||||||||||:::|||   :::   |||   |||:::||||||
Db        617 pPheValSerThrProProLeuValArgLeuValPheAsnLeuValLysAspProIlePr 637

Qy       1965 CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGC 2024
               ||||||||||||||||  ||||||:::  ||||||::::::|||   ||||||||
Db        637 oGlyAlaGluThrPheTyrThrAspGlySerCysAsnArgGlnSerLysGluGlyLysAl 657

Qy       2025 CGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCA 2084
              |||||||||||||||:::||||||  |||:::       |||   :::|||||||||||
Db        657 aGlyTyrValThrAspLysGlyArgAspLysValLysValLeuGluGlnThrThrAsnGl 677

Qy       2085 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACAT 2144
              |:::   ||||||:::|||      :::||||||  |||||||||   :::   |||||
Db        677 nGlnAlaGluLeuGluAlaPheAlaMetAlaLeuThrAspSerGlyProLysAlaAsnIl 697

Qy       2145 CGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAG 2204
              |:::   ||||||||||||      :::||||||:::       ||||||   :::||||||::
Db        697 eIleValAspSerGlnTyrValMetGlyIleValValGlyGlnProThrGluSerGluAs 717

Qy       2205 CGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTG 2264
              :    :::||||||||||||||||||::::::||||||||||||   :::|||:::::::||
Db        717 nArgIleValAsnGlnIleIleGluGluMetIleLysLysGluAlaIleTyrValAlaTr 737

Qy       2265 GGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2324
              ||||||||||||||||||||||||||||||:::::::::|||   |||||||||:::||
Db        737 pValProAlaHisLysGlyIleGlyGlyAsnGlnGluValAspHisLeuValSerGlnGl 757

Qy       2325 CATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||||||:::||||||||||||:::   |||:::
Db        757 yIleArgGlnValLeuPheLeuGluLysIleGlu 768
```

Search completed: June 2, 2005, 05:09:55
Job time : 107.734 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  1, 2005, 20:20:55 ; Search time 208.158 Seconds
                                          (without alignments)
                                          12088.668 Million cell updates/sec

Title:          US-09-610-313B-32
Perfect score:  4555
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2457

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters:      3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155136_15147/app_query.fasta_1.7893
-DB=UniProt_03 -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09610313_@CGN_1_1_725_@runat_31052005_155136_15147 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      UniProt_03:*
                1:  uniprot_sprot:*
                2:  uniprot_trembl:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4097	89.9	1427	2	Q9WF92	Q9wf92 human immun
2	4093	89.9	1427	2	Q9WF89	Q9wf89 human immun
3	4089	89.8	1428	2	Q9WF86	Q9wf86 human immun
4	3947	86.7	1427	2	Q9WF62	Q9wf62 human immun
5	3940	86.5	1429	2	Q6X4P8	Q6x4p8 human immun
6	3912	85.9	1425	2	Q6X4Q6	Q6x4q6 human immun
7	3905	85.7	1427	2	Q9WF96	Q9wf96 human immun
8	3901	85.6	1430	2	Q6X4R4	Q6x4r4 human immun
9	3892.5	85.5	1427	2	Q9WF71	Q9wf71 human immun
10	3888	85.4	1427	2	Q9WFA0	Q9wfa0 human immun
11	3885.5	85.3	1427	2	Q9WF65	Q9wf65 human immun
12	3884	85.3	1437	2	Q9WF77	Q9wf77 human immun
13	3876.5	85.1	1427	2	Q9WF74	Q9wf74 human immun
14	3870.5	85.0	1427	2	Q8ADX5	Q8adx5 human immun
15	3862	84.8	1425	2	Q9WF79	Q9wf79 human immun
16	3859	84.7	1427	2	Q9WF98	Q9wf98 human immun
17	3850.5	84.5	1429	2	Q9WF84	Q9wf84 human immun
18	3841	84.3	1428	2	Q6X4P0	Q6x4p0 human immun
19	3831.5	84.1	1427	2	Q9WF68	Q9wf68 human immun
20	3795.5	83.3	1432	2	Q8ADZ9	Q8adz9 human immun
21	3792	83.2	1433	2	Q8ADE9	Q8ade9 human immun
22	3790	83.2	1434	2	Q8ADV1	Q8adv1 human immun
23	3788	83.2	1430	2	Q7SV36	Q7sv36 human immun
24	3784	83.1	1425	2	Q7SV20	Q7sv20 human immun
25	3778.5	83.0	1433	2	Q8ADN7	Q8adn7 human immun
26	3777.5	82.9	1433	2	Q8ADK5	Q8adk5 human immun
27	3775.5	82.9	1433	2	Q8ADS7	Q8ads7 human immun
28	3772.5	82.8	1433	2	Q8ADG5	Q8adg5 human immun
29	3769.5	82.8	1435	2	Q9DQ33	Q9dq33 human immun
30	3768.5	82.7	1433	2	Q8ADE1	Q8ade1 human immun
31	3763.5	82.6	1431	2	Q8ADR1	Q8adr1 human immun
32	3762.5	82.6	1433	2	Q8ADC5	Q8adc5 human immun
33	3761.5	82.6	1433	2	Q8AE32	Q8ae32 human immun
34	3760.5	82.6	1433	2	Q8ADM1	Q8adm1 human immun
35	3760.5	82.6	1433	2	Q8ADU3	Q8adu3 human immun
36	3759.5	82.5	1437	2	O41782	O41782 human immun
37	3758	82.5	1434	2	Q8Q7Z7	Q8q7z7 human immun
38	3758	82.5	1435	2	Q9WLJ3	Q9wlj3 human immun
39	3757.5	82.5	1433	2	Q8AD85	Q8ad85 human immun
40	3755.5	82.4	1433	2	Q7SPT9	Q7spt9 human immun
41	3751.5	82.4	1433	2	Q8ADW7	Q8adw7 human immun
42	3751.5	82.4	1436	2	Q8AD93	Q8ad93 human immun
43	3749.5	82.3	1433	2	Q8AE23	Q8ae23 human immun
44	3748.5	82.3	1433	2	Q8ADR9	Q8adr9 human immun
45	3746.5	82.3	1433	2	Q8ADH3	Q8adh3 human immun

ALIGNMENTS

```
RESULT 1
Q9WF92
ID   Q9WF92      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF92;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110975; AAD17135.1; -.
DR   HSSP; P04585; 1JLA.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
```

```
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW   Protease; RNA-directed DNA polymerase; Transferase.
SQ   SEQUENCE   1427 AA;  161059 MW;  F3B0C741A1EFB10A CRC64;

Alignment Scores:
Pred. No.:              3.67e-155     Length:        1427
Score:                  4097.00       Matches:       777
Percent Similarity:     98.98%        Conservative:  1
Best Local Similarity:  98.85%        Mismatches:    4
Query Match:            89.95%        Indels:        4
DB:                     2             Gaps:          2

US-09-610-313B-32 (1-2457) x Q9WF92 (1-1427)

Qy          13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
               :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379

Qy          73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy         133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy         193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439

Qy         253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459

Qy         313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479

Qy         373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy         433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519

Qy         493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy         553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559

Qy         613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy         673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy         733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy         793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
               |||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db         620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy         853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy         913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy         973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy        1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy        1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739

Qy        1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759

Qy        1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
               ||||||      ||||||||||||||||||||||||||||||||||||||||||||||||
Db         760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy        1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy        1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
               ||||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db         800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy        1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy        1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy        1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139

Qy       2341 TTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 2
Q9WF89
ID   Q9WF89      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF89;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110974; AAD17126.1; -.
DR   HSSP; P04585; 1JLA.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  161293 MW;  270F25DF67615C06 CRC64;

Alignment Scores:
Pred. No.:              5.29e-155     Length:         1427
Score:                  4093.00       Matches:        776
Percent Similarity:     98.85%        Conservative:   1
Best Local Similarity:  98.73%        Mismatches:     5
Query Match:            89.86%        Indels:         4
DB:                     2             Gaps:           2

US-09-610-313B-32 (1-2457) x Q9WF89 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
```

```
                :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||            ||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||            ||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||     |||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099
```

```
Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139

Qy       2341 TTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 3
Q9WF86
ID   Q9WF86      PRELIMINARY;      PRT;  1428 AA.
AC   Q9WF86;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110973; AAD17119.1; -.
DR   HSSP; P04585; 1JLA.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1428 AA;  161328 MW;  2D61CD9FEABD2387 CRC64;

Alignment Scores:
Pred. No.:               7.63e-155     Length:          1428
Score:                   4089.00       Matches:         775
Percent Similarity:      98.98%        Conservative:    3
Best Local Similarity:   98.60%        Mismatches:      4
Query Match:             89.77%        Indels:          4
DB:                      2             Gaps:            2

US-09-610-313B-32 (1-2457) x Q9WF86 (1-1428)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetMetGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||            ||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              |||||||||||||||||||||||             ||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAspArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139

Qy       2341 TTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 4
Q9WF62
ID   Q9WF62      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF62;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110961; AAD17021.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
```

```
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  160992 MW;  AC2ACBF190C77DAD CRC64;

Alignment Scores:
Pred. No.:               3.34e-149     Length:        1427
Score:                   3947.00       Matches:       742
Percent Similarity:      96.82%        Conservative:  19
Best Local Similarity:   94.40%        Mismatches:    21
Query Match:             86.65%        Indels:        4
DB:                      2             Gaps:          2

US-09-610-313B-32 (1-2457) x Q9WF62 (1-1427)

Qy        13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
             :::|||||||||||  ||||||  |||    ||||||:::||||||||||||||||||
Db       360 LeuAlaGluAlaMetGlyGlnAlaAsnSerMetAsnIleMetMetGlnArgSerAsnPhe 379

Qy        73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
             |||   |||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db       380 LysAsnProLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy       133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy       193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
             |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db       420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy       253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
             ||||||||||||||||||||||||||||   ||||||    ||||||||||||||||||||
Db       440 LysAlaArgGluPheProSerGluGlnThrArgAlaIleSerProThrSerArgGluLeu 459

Qy       313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
             ||||||||||||||||||||||||||||||||||||||||      ||||||    ||||||
Db       460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluGlyGlnGlyAlaLeuAsn 479

Qy       373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
                |||||||||||||||||||||||||||||||:::|||:::||||||||||||||||
Db       480 LeuProGlnIleThrLeuTrpGlnArgProLeuValThrIleArgValGlyGlyGlnIle 499

Qy       433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
             ||||||||||||||||||||||||||||||||||||||||||||||||:::::||||||
Db       500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519

Qy       493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy       553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
             |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy       613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy       673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy       733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy       793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
             ||||||    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       620 MetGluMetGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy       853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy       913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy       973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy      1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
             ||||||    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       700 AspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy      1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
             |||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy      1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
             ||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db       740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaLeuAsnProGluIleValIle 759

Qy      1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
             ||||||      ||||||||||||||||||||||||||||||||||||||||||||||||
Db       760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy      1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
             |||||||||||||||:::|||||||||:::||||||||||||||||||||||||||||||
Db       780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799

Qy      1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
             ||||||||||||||||||||||||            ||||||||||||||||||||||||
Db       800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819
```

```
Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        820 ValGlnProIleAsnLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIle 959

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||:::|||||||||:::|||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluGluGluProIleValGlyAlaGluThrPhe 1019

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ArgGlyArgGlnArgIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              |||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGlyValAsnIleValThrAspSerGln 1079

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||   ||||||:::|||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyLysGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2341 TTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 5
Q6X4P8
ID   Q6X4P8      PRELIMINARY;      PRT;  1429 AA.
AC   Q6X4P8;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14709249; DOI=10.1089/088922203771881220;
RA   Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,
RA   McCutchan F.E.;
RT   "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT   in full genome analysis, is not sustained in subgenomic regions.";
RL   AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., McCutchan F.E.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maayan S., Zeira M.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AY255825; AAP76538.1; -.
DR   HSSP; P03367; 1A8G.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
```

```
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   1429 AA;  161586 MW;  D0B9CD124DC64E19 CRC64;

Alignment Scores:
Pred. No.:              6.34e-149      Length:         1429
Score:                  3940.00        Matches:        739
Percent Similarity:     96.96%         Conservative:   27
Best Local Similarity:  93.54%         Mismatches:     16
Query Match:            86.50%         Indels:         8
DB:                     2              Gaps:           3

US-09-610-313B-32 (1-2457) x Q6X4P8 (1-1429)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||   :::||||||||:::||||||:::|||||||||
Db        358 LeuAlaGluAlaMetSerGlnAlaAsnAsnAlaAsnIleMetMetGlnLysSerAsnPhe 377

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        378 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 397

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        398 AsnCysArgAlaProArgLysArgGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 417

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              |||||||||||||||||||||||||||||||||||||||:::|||||||||    ||||||
Db        418 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheGlnGlnGly 437

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db        438 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 457

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGC--------- 363
              ||||||||||||||||||||    ||||||||||||    |||||||||||||
Db        458 GlnValArgGlyAspAsnProIleSerGluAlaGlyAspGluArgGlnGlyAlaLeuGln 477

Qy        364 ---ACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
                 |||||||||    ||||||||||||||||||||||||||||:::||||||:::
Db        478 ArgThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIle 497

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGlu 517

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ::::::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 IleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 537

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        538 ValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrVal 557

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 LeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCys 577

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 597

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              |||||||||||||||||||||||||||||||:::|||||||||:::||||||||||||
Db        598 AspGlyProLysValLysGlnTrpProLeuThrLysGluLysIleGluAlaLeuThrAla 617

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 IleCysGluGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyr 637

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 AsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAsp 657

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 677

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 ProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 697

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db        698 SerValProLeuAspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsn 717

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 737

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              |||:::|||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        738 ProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnPro 757

Qy       1201 GAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              |||:::|||||||||||          |||||||||||||||||||||||||||||
Db        758 GluLeuValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 777

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        778 GlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThr 797

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCC 1368
              ||||||||||||||||||||||||||||||||||||          ||||||||||||
Db        798 ProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro 817

Qy       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        818 AspLysTrpThrValGlnProIleGlnLeuProGluLysGluSerTrpThrValAsnAsp 837

Qy       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal 857

Qy       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThr 877

Qy       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
              ||||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db        878 GluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGly 897

Qy       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        898 ValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGln 917

Qy       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 937

Qy       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 MetArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 957

Qy       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
              |||||    |||||||||||||||||    ||||||||||||||||||||||||||||||
Db        958 MetGluGlyIleValIleTrpGlyLysValProLysPheArgLeuProIleGlnLysGlu 977

Qy       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db         978 ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProAspTrpGluPhe 997

Qy        1909 GTGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db         998 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGly 1017

Qy        1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1018 ValGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1037

Qy        2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1038 TyrValThrAspArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLys 1057

Qy        2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1058 ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1077

Qy        2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1078 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 1097

Qy        2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
               ||||||||||||||||||||||||||||||:::||||||:::||||||||||||||||||
Db        1098 LeuValAsnGlnIleIleGluGlnLeuIleArgLysGluArgValTyrLeuSerTrpVal 1117

Qy        2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
               ||||||||||||||||||||||||||||||:::|||||||||||||||    ||||||
Db        1118 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIle 1137

Qy        2329 CGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
               ||||||||||||||||||||||||||||||
Db        1138 ArgLysValLeuPheLeuAspGlyIleAsp 1147
```

```
RESULT 6
Q6X4Q6
ID   Q6X4Q6      PRELIMINARY;      PRT;  1425 AA.
AC   Q6X4Q6;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14709249; DOI=10.1089/088922203771881220;
RA   Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,
RA   McCutchan F.E.;
RT   "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT   in full genome analysis, is not sustained in subgenomic regions.";
RL   AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., McCutchan F.E.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maayan S., Zeira M.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AY255824; AAP76529.1; -.
DR   HSSP; Q72498; 1K6Y.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   1425 AA;  161115 MW;  782315D4D5305E18 CRC64;

Alignment Scores:
Pred. No.:              8.2e-148      Length:         1425
Score:                  3912.00       Matches:        730
Percent Similarity:     96.69%        Conservative:   30
Best Local Similarity:  92.88%        Mismatches:     22
Query Match:            85.88%        Indels:         4
DB:                     2             Gaps:           2

US-09-610-313B-32 (1-2457) x Q6X4Q6 (1-1425)

Qy          13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
               :::||||||||||:::||||||   :::   :::|||:::|||||||||:::||||||
Db         358 LeuAlaGluAlaMetThrGlnAlaAsnAsnThrHisIleMetMetGlnArgAsnAsnPhe 377

Qy          73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
               ||||||||||||||||:::||||||||||||||||||||||||||||||:::||||||
Db         378 LysGlyProLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 397

Qy         133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         398 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 417

Qy         193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
```

```
Db     418 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluIleLeuAlaPheProGlnGly 437

Qy     253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
Db     438 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrAsnGlyGluLeu 457

Qy     313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
Db     458 GlnValArgArgAsnAsnProArgSerLysThrGlyValGluArgGlnGlyThrLeuAsn 477

Qy     373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
Db     478 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 497

Qy     433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
Db     498 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 517

Qy     493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
Db     518 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 537

Qy     553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
Db     538 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 557

Qy     613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
Db     558 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 577

Qy     673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
Db     578 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 597

Qy     733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
Db     598 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 617

Qy     793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db     618 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 637

Qy     853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db     638 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 657

Qy     913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db     658 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 677

Qy     973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db     678 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 697

Qy    1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db     698 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 717

Qy    1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db     718 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 737

Qy    1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db     738 GlnSerSerMetThrArgIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 757

Qy    1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db     758 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 777

Qy    1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db     778 LysIleGluGluLeuArgGlyHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 797

Qy    1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
Db     798 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 817

Qy    1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
Db     818 ValGlnProIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 837

Qy    1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
Db     838 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 857

Qy    1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
Db     858 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 877

Qy    1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
Db     878 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 897

Qy    1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
Db     898 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyAspAspGlnTrpThrTyrGln 917

Qy    1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
Db     918 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 937

Qy    1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
Db     938 HisThrAsnAspValLysGlnLeuValGlyAlaValGlnLysIleAlaMetGluGlyIle 957

Qy    1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
Db     958 ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 977

Qy    1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
Db     978 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProAspTrpGluPheValAsnThrPro 997

Qy    1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
Db     998 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyValGluThrPhe 1017

Qy    1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
Db    1018 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1037

Qy    2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
Db    1038 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1057

Qy    2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
Db    1058 AlaIleGlnIleAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1077

Qy    2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
Db    1078 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1097

Qy    2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
Db    1098 IleIleGluGlnLeuIleLysLysGluArgIleTyrLeuThrTrpValProAlaHisLys 1117

Qy    2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
Db    1118 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAsnGlyIleArgLysValLeu 1137

Qy    2341 TTCCTGGACGGCATCGAT 2358
Db    1138 PheLeuAspGlyIleAsp 1143
```

```
RESULT 7
Q9WF96
ID   Q9WF96      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF96;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110976; AAD17144.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  160518 MW;  FBDA73D603A68B18 CRC64;

Alignment Scores:
Pred. No.:              1.56e-147      Length:        1427
Score:                  3905.00        Matches:       726
Percent Similarity:     96.56%         Conservative:  33
Best Local Similarity:  92.37%         Mismatches:    23
Query Match:            85.73%         Indels:        4
DB:                     2              Gaps:          2

US-09-610-313B-32 (1-2457) x Q9WF96 (1-1427)

Qy          13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
               :::|||||||||||||||||||||   :::   ||||||::::::|||||||||||||||
Db         360 LeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetIleGlnArgSerAsnPhe 379

Qy          73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
               |||||||||:::|||   :::|||||||||||||||||||||||||||||||||||||||
Db         380 LysGlyProArgArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy         133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
               ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db         400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet 419

Qy         193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
               |||||||||||||||||||||||||||||||||   |||:::||||||||||||||||||
Db         420 LysAspCysThrGluArgGlnAlaAsnPhePheGlyGluAsnLeuAlaPheProGlnGly 439

Qy         253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
               :::||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db         440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy         313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
               :::|||   |||||||||||||||||||||   ||||||:::   |||||||||   |||
Db         460 ArgValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479

Qy         373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
               ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||:::
Db         480 PheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysValGlyGlyGlnVal 499

Qy         433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
               ||||||||||||||||||||||||||||||||||||||||||||||||::::::||||||
Db         500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519

Qy         493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy         553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
               |||||||||   ||||||||||||||||||||||||||||||||||||||:::||||||
Db         540 AspGlnIleThrIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy         613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
               |||||||||||||||||||||||||||:::|||||||||:::||||||||||||||||||
Db         560 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe 579

Qy         673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy         733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy         793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy         853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   ||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        700 AspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              |||||||||||||||||||||||||||||||||||||||:::||||||   |||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||            ||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||   |||||||||:::|||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||            ||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||:::||||||:::|||:::||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||   ||||||||||||||||||   ||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db       1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||   ||||||::::::|||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              |||||||||||||||||||||:::|||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2341 TTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 8
Q6X4R4
ID   Q6X4R4      PRELIMINARY;      PRT;  1430 AA.
AC   Q6X4R4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14709249; DOI=10.1089/088922203771881220;
RA   Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,
RA   McCutchan F.E.;
RT   "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT   in full genome analysis, is not sustained in subgenomic regions.";
RL   AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., McCutchan F.E.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maayan S., Zeira M.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AY255823; AAP76521.1; -.
DR   HSSP; P12497; 1B9D.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
```

```
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   1430 AA;  161540 MW;  5BBA39B85D03FA0B CRC64;

Alignment Scores:
Pred. No.:              2.24e-147      Length:        1430
Score:                  3901.00        Matches:       727
Percent Similarity:     96.58%         Conservative:  36
Best Local Similarity:  92.03%         Mismatches:    19
Query Match:            85.64%         Indels:        8
DB:                     2              Gaps:          3

US-09-610-313B-32 (1-2457) x Q6X4R4 (1-1430)

Qy       13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
            :::|||||||||||||||||   |||:::   ||||||:::||||||:::|||||||||
Db      359 LeuAlaGluAlaMetSerGlnValThrAsnThrAsnIleMetMetGlnLysSerAsnPhe 378

Qy       73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
            |||||||||::::::   ||||||||||||||||||||||||||||||||:::||||||
Db      379 LysGlyProArgLysThrIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 398

Qy      133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
            ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db      399 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyArgGluGlyHisGlnMet 418

Qy      193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
            |||||||||||||||||||||||||||||||||||||||:::|||||||||      ||||||
Db      419 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheGlnGlnGly 438

Qy      253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
            :::||||||||||||||   ||||||   ||||||||||||||||||||||||||||||
Db      439 GluAlaArgGluPheProProGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 458

Qy      313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC------------CAG 360
            |||||||||||||||||||   ||||||||||||   ||||||              |||
Db      459 GlnValArgGlyAspAsnProLeuSerGluAlaGlyThrGluArgProGlyThrLeuGln 478

Qy      361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
            ||||||||||||   |||||||||||||||||||||||||||||||:::||||||:::
Db      479 GlyThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIle 498

Qy      421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
            ||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db      499 GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGlu 518

Qy      481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
            ::::::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      519 IleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 538

Qy      541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db      539 ValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlySerVal 558

Qy      601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
            |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      559 LeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCys 578

Qy      661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
            ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db      579 ThrLeuAsnPheProIleSerProIleGluThrIleProValLysLeuLysProGlyMet 598

Qy      721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      599 AspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAla 618

Qy      781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
            ||||||:::||||||:::||||||||||||||||||||||||||||||||||||||||||
Db      619 IleCysGluAspMetGluArgGluGlyLysIleSerLysIleGlyProGluAsnProTyr 638

Qy      841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      639 AsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAsp 658

Qy      901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      659 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 678

Qy      961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      679 ProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 698

Qy     1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
            ||||||||||||||||||:::|||||||||||||||||||||||||||||||||   |||
Db      699 SerValProLeuAspGluAsnPheArgLysTyrThrAlaPheThrIleProSerThrAsn 718

Qy     1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      719 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 738

Qy     1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
            |||:::||||||||||||||||||||||||||||||||||||||||||:::||||||
Db      739 ProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnPro 758

Qy     1201 GAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
            ||||||||||||||||||                ||||||||||||||||||||||||||
Db      759 GluIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 778

Qy     1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
            ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db      779 GlnHisArgAlaLysValGluGluLeuArgGlyHisLeuLeuArgTrpGlyPheThrThr 798

Qy     1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCC 1368
            ||||||||||||||||||||||||||||||||||||            ||||||||||||
Db      799 ProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro 818
```

```
Qy       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
              ||||||||||||||||||   |||:::||||||||||||:::||||||||||||||||||
Db        819 AspLysTrpThrValGlnThrIleGlnLeuProGluLysAspSerTrpThrValAsnAsp 838

Qy       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal 858

Qy       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
              |||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        859 ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValAlaLeuThr 878

Qy       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        879 GluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGly 898

Qy       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        899 ValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGln 918

Qy       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        919 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 938

Qy       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        939 ArgArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 958

Qy       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 MetGluSerIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGlu 978

Qy       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        979 ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 998

Qy       1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        999 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 1018

Qy       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1038

Qy       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
              ||||||||||||:::||||||||||||:::|||||||||   ||||||||||||||||||
Db       1039 TyrValThrAspLysGlyArgGlnLysValValSerLeuAsnGluThrThrAsnGlnLys 1058

Qy       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1059 ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1078

Qy       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       1079 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerAspSerGlu 1098

Qy       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
              :::||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db       1099 IleValAsnGlnIleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpVal 1118

Qy       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db       1119 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerLysGlyIle 1138

Qy       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||||||||||||||
Db       1139 ArgLysValLeuPheLeuAspGlyIleAsp 1148

RESULT 9
Q9WF71
ID   Q9WF71      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF71;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110964; AAD17047.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  161326 MW;  B9FC43C29B26CFA2 CRC64;

Alignment Scores:
```

```
Pred. No.:               4.88e-147     Length:         1427
Score:                   3892.50       Matches:        723
Percent Similarity:      96.44%        Conservative:   35
Best Local Similarity:   91.98%        Mismatches:     23
Query Match:             85.46%        Indels:         5
DB:                      2             Gaps:           3

US-09-610-313B-32 (1-2457) x Q9WF71 (1-1427)

Qy       13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
            :::||||||||||||||||||   |||:::   |||::::::||||||||||||||||||
Db      361 LeuAlaGluAlaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsnPhe 379

Qy       73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
            |||||||||:::||||||:::|||||||||||||||||||||||||||||||||||||||
Db      380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy      133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy      193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
            |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db      420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy      253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
            :::||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db      440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy      313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
            ||||||||||||||||||||||||||||||      ||||||   ||||||      |||
Db      460 GlnValArgGlyAspAsnProArgSerGluThrArgAlaGluGlyGlnGlyAsnPheAsn 479

Qy      373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy      433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
            :::||||||||||||||||||||||||||||||||||||||||:::::::::||||||
Db      500 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519

Qy      493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
            ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db      520 GlyLysTrpLysProLysIleIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy      553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
            :::||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db      540 GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy      613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
            |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      560 ThrProIleAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy      673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy      733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy      793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy      853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy      913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy      973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy     1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
            ||||||   ||||||||||||||||||||||||||||||||||||||||||   ||||||
Db      700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719

Qy     1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
            |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db      720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy     1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCCGCAACCCCGAGATCGTGATC 1212
            ||||||||||||||||||||||||||||||||||||   :::||||||:::|||||||||
Db      740 GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759

Qy     1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
            ||||||            ||||||||||||:::|||||||||||||||||||||||||||
Db      760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy     1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
            ||||||||||||||||||||||||||||||:::|||:::|||   |||||||||||||||
Db      780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys 799

Qy     1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
            ||||||||||||||||||||||||            ||||||||||||||||||||||||
Db      800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy     1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
            ||||||||||||:::||||||:::|||:::||||||||||||||||||||||||||||||
Db      820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy     1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
            :::||||||||||||||||||||||||||||||||||||:::|||:::      ||||||
Db      840 IleGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysHisLeuCys 859

Qy     1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy     1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
            |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db      880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy     1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
            ||||||||||||:::||||||:::||||||||||||||||||||||||||||||||||||
Db      900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy     1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy     1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
            |||||||||||||||||||||||||||   ||||||||||||||||||||||||   |||
Db      940 HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleAlaMetGluCysIle 959

Qy     1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
            |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db      960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979

Qy     1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy     1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
            ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db     1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019

Qy     1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
            ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db     1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039

Qy     2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
```

```
                ::|||||||||||||||||||||||||   ||||||||||||||||||   |||||||||
Db       1040 LysGlyArgGlnLysIleValSerLeuAsnGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||||:::||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2341 TTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 10
Q9WFA0
ID   Q9WFA0      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WFA0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110978; AAD17162.1; -.
DR   PIR; S49086; S49086.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  160503 MW;  24C410797550C0F8 CRC64;

Alignment Scores:
Pred. No.:              7.37e-147     Length:        1427
Score:                  3888.00       Matches:       723
Percent Similarity:     96.44%        Conservative:  35
Best Local Similarity:  91.98%        Mismatches:    24
Query Match:            85.36%        Indels:        4
DB:                     2             Gaps:          2

US-09-610-313B-32 (1-2457) x Q9WFA0 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||    :::   ||||||:::::::||||||||||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetIleGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::|||    :::|||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||:::||||||||||||:::||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAspPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||    ||||||||||||||||||||   ||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerGlyGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||   ||||||||||||||||||||||   ||||||:::   |||||||||   |||
Db        460 GlnValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnVal 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||::::::|||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539
```

```
Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
Db        540 AspGlnIleThrIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
Db        560 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
Db        700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db        780 LysIleGluGluLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
Db        820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
Db        900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
Db        940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
Db        960 ValLeuTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProLysTrpGluPheValAsnThrPro 999

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyCysValThrAsp 1039

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
Db       1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
Db       1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
Db       1120 GlyIleGlyGlyHisGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2341 TTCCTGGACGGCATCGAT 2358
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 11
Q9WF65
ID   Q9WF65      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF65;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
```

```
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110962; AAD17030.1; -.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW   Protease; RNA-directed DNA polymerase; Transferase.
SQ   SEQUENCE   1427 AA;  161347 MW;  595F19865DE81D63 CRC64;

Alignment Scores:
Pred. No.:              9.26e-147     Length:          1427
Score:                  3885.50       Matches:         724
Percent Similarity:     96.44%        Conservative:    34
Best Local Similarity:  92.11%        Mismatches:      23
Query Match:            85.30%        Indels:          5
DB:                     2             Gaps:            3

US-09-610-313B-32 (1-2457) x Q9WF65 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
                 :::||||||||||||||||   |||:::   |||::::::|||||||||||||||||
Db        361 LeuAlaGluAlaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::|||||:::||||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::|||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||||||||        ||||||    |||||||||||||
Db        460 GlnValArgGlyAspAsnProArgSerGluThrArgAlaGluGlyGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              :::||||||||||||||||||||||||||||||||||||||||:::::::::||||||
Db        500 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              :::|||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        540 GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   |||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||   |||||||||||||||||||||   :::||||||:::||||||||||
Db        740 GlnSerSerMetIleLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||            ||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779
```

```
Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||:::|||||||||:::||||||   |||||||||   ||||||
Db        780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProTyrLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||:::||||||:::|||:::|||||||||||:::||||||||||||||
Db        820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValHisAspIleGlnLysLeu 839

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||:::|||:::   ||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysHisLeuCys 859

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||:::||||||:::||||||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||   ||||||||||||   ||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleThrMetGluSerIle 959

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              :::||||||||||||||||   ||||||||||||||||||||:::   |||||||||
Db       1040 LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnArgAlaGluLeuGln 1059

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||:::||||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2341 TTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 12
Q9WF77
ID   Q9WF77      PRELIMINARY;      PRT;  1437 AA.
AC   Q9WF77;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110967; AAD17072.1; -.
DR   HSSP; P24736; 1NCP.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
```

```
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1437 AA;  161978 MW;  65DAA2562FC317A6 CRC64;

Alignment Scores:
Pred. No.:               1.06e-146      Length:         1437
Score:                   3884.00        Matches:        732
Percent Similarity:      95.73%         Conservative:   30
Best Local Similarity:   91.96%         Mismatches:     20
Query Match:             85.27%         Indels:         14
DB:                      2              Gaps:           4

US-09-610-313B-32 (1-2457) x Q9WF77 (1-1437)

Qy        13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
             :::||||||||||||||||||   |||   |||||:::||||||:::||||||||
Db       360 LeuAlaGluAlaMetSerGlnAlaAsnSerValAsnIleMetMetGlnLysSerAsnPhe 379

Qy        73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
             |||||||||:::|||   :::||||||||||||||||||||||||||||||||||:::
Db       380 LysGlyProArgArgAsnValLysCysPheAsnCysGlyLysGluGlyHisIleAlaLys 399

Qy       133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy       193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
             ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db       420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy       253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCC------------------AACAGC 294
             :::||||||||||||||    ||||||   ||||||                  ||||||
Db       440 GluAlaArgGluPheProProGluGlnIleArgAlaSerSerProAsnSerThrAsnSer 459

Qy       295 CCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAG 354
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       460 ProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGlu 479

Qy       355 ------------CGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
                         |||||||||||||||   ||||||||||||||||||||||||||
Db       480 GlyGlnGlyThrLeuGlnGlyThrLeuAsnCysProGlnIleThrLeuTrpGlnArgPro 499

Qy       403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       500 LeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAlaAsp 519

Qy       463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
             :::|||||||||||||||::::::||||||||||||||||||||||||||||||||||||
Db       520 AsnThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 539

Qy       523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
             |||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db       540 IleGlyGlyPheIleLysValArgGlnTyrAspGlnIleValIleGluIleCysGlyLys 559

Qy       583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
             |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db       560 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 579

Qy       643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       580 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 599

Qy       703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
             ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db       600 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpThrLeuThrGluGluLys 619

Qy       763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
             |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db       620 IleLysAlaLeuThrGluIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 639

Qy       823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       640 GlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLys 659

Qy       883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       660 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 679

Qy       943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       680 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp 699

Qy      1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
             :::||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db       700 MetGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 719

Qy      1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       720 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 739

Qy      1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
             ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db       740 GlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluPro 759

Qy      1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGC 1236
             ||||||   :::|||||||||||||||||||||||||            ||||||||||||
Db       760 PheArgLeuGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 779

Qy      1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
             |||||||||||||||||||||   ||||||:::||||||||||||||||:::||||||||
Db       780 SerAspLeuGluIleGlyGlnArgArgAlaGlnIleGluGluLeuArgGluHisLeuLeu 799

Qy      1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------ 1350
             :::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db       800 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 819

Qy      1351 CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1410
                   |||||||||||||||||||||||||||||||:::||||||:::|||:::
Db       820 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProAspLysAsp 839

Qy      1411 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1470
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       840 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 859

Qy      1471 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1530
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       860 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThr 879

Qy      1531 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1590
             |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       880 AspValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu 899

Qy      1591 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1650
             :::|||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db       900 LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln 919

Qy      1651 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1710
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       920 LysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 939

Qy      1711 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1770
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       940 ThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThrGlu 959

Qy      1771 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1830
             ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db       960 AlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThrProLysPheArg 979

Qy      1831 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1890
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       980 LeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrp 999
```

```
Qy       1891 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 1019

Qy       1951 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2010
              ||||||||||||      |||   ||||||||||||||||||||||||||||||||||||
Db       1020 LysGluProIleProGlyValGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 1039

Qy       2011 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2070
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 LysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSerLeuThr 1059

Qy       2071 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGly 1079

Qy       2131 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 SerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 1099

Qy       2191 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2250
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       1100 AspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleGlnLysGluTrp 1119

Qy       2251 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2310
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       1120 ValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLys 1139

Qy       2311 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||||||||:::|||||||||:::||||||||||||||||||||||||
Db       1140 LeuValSerGlnGlyIleArgLysMetLeuPheLeuAspGlyIleAsp 1155

RESULT 13
Q9WF74
ID   Q9WF74      PRELIMINARY;      PRT;  1427 AA.
AC   Q9WF74;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110965; AAD17056.1; -.
DR   PIR; S49086; S49086.
DR   HSSP; Q70622; 1HVN.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1427 AA;  161299 MW;  924682CF5095D763 CRC64;

Alignment Scores:
Pred. No.:              2.11e-146     Length:         1427
Score:                  3876.50       Matches:        720
Percent Similarity:     96.31%        Conservative:   37
Best Local Similarity:  91.60%        Mismatches:     24
Query Match:            85.10%        Indels:         5
DB:                     2             Gaps:           3

US-09-610-313B-32 (1-2457) x Q9WF74 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::|||||||||||||||||||||     :::   :::::::::||||||:::|||||||||
Db        361 LeuAlaGluAlaMetSerGlnAla---AsnThrSerValMetMetGlnLysSerAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||:::||||||:::|||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              |||||||||||||||||||||||||||            |||    ||||||      |||
Db        460 GlnValArgGlyAspAsnProArgSerGluThrArgValGluGlyGlnGlyAsnPheAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              :::||||||||||||||||||||||||||||||||||||||||||:::::::::||||||
Db        500 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              :::||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        540 GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnPheGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp 619

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   ||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              ||||||||||||||||||||||||||||||||||||   :::||||||:::|||||||||
Db        740 GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||            ||||||||||||||||||||||||||||||||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||:::|||||||||:::||||||   ||||||||||||||||||
Db        780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||            ||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||:::||||||:::|||:::||||||||||||||||||||||||||||||
Db        820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||:::||||||   ||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValArgHisLeuCys 859

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              |||||||||||||||:::||||||:::|||||||||||||||||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||   |||:::|||||||||||||||||||||
Db        940 HisThrAsnAspValLysGlnLeuThrGluValValArgLysIleAlaMetGluSerIle 959

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              |||||||||||||||||||||||||||||||||:::|||||||||||||||||||||:::
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 1039

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              :::||||||||||||||||||   ||||||||||||||||||||||||   |||||||||
Db       1040 LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              |||||||||||||||||||||||||||::::::|||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              |||||||||||||||||||||:::|||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2341 TTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145
```

```
RESULT 14
Q8ADX5
ID   Q8ADX5      PRELIMINARY;      PRT;  1427 AA.
AC   Q8ADX5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol fusion polyprotein (Fragment).
GN   Name=gag-pol;
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22375625; PubMed=12487816; DOI=10.1089/088922202320886325;
RA   Harris M.E., Serwadda D., Sewankambo N., Wabwire F., Kim B.,
RA   Kigozi G., Kiwanuka N., Phillips J.B., Meehen M., Lutalo T.,
RA   Lane J.R., Merling R., Gray R., Wawer M., Birx D.L., Robb M.L.,
RA   McCutchan F.E.;
RT   "Among 46 near full length HIV type 1 genome sequences from Rakai
RT   District, Uganda, subtype D and AD recombinants predominate.";
RL   AIDS Res. Hum. Retroviruses 18:1281-1290(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Harris M.E., Birx D.L., Robb M.L.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kim B., Phillips J.B., Lane J.R., Merling R., McCutchan F.E.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Lutalo T.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Meehen M., Wawer M.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Serwadda S., Sewankambo N., Wabwire F., Kigozi G., Kiwanuka N.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF484491; AAN73556.1; -.
DR   HSSP; P24736; 1NCP.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW   Protease; RNA-directed DNA polymerase; Transferase.
FT   NON_TER       1      1
SQ   SEQUENCE   1427 AA;  161058 MW;  20559041E2A88B5C CRC64;

Alignment Scores:
Pred. No.:              3.65e-146     Length:         1427
Score:                  3870.50       Matches:        720
Percent Similarity:     96.18%        Conservative:   36
Best Local Similarity:  91.60%        Mismatches:     25
Query Match:            84.97%        Indels:         5
DB:                     2             Gaps:           3

US-09-610-313B-32 (1-2457) x Q8ADX5 (1-1427)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||         :::||||||:::|||||||:::         |||
Db        361 LeuAlaGluAlaMetSerGlnValGlnHisSerAsnIleMetMetGlnGlnGlyLysPhe 380

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              :::|||   ||||||   ||||||||||||||||||||||||||||||||||||:::||||||
Db        381 ArgGlyGlnLysArg---IleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
              ||||||||||||||||||||||||   ||||||||||||||||   ||||||||||||||||
Db        460 GlnValArgGlyAspAsnProArgPheGluAlaGlyAlaGluGlyGlnGlyThrLeuAsn 479

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              |||||||||||||||||||||||||||||   |||:::|||:::|||||||||||||:::
Db        480 PheProGlnIleThrLeuTrpGlnArgProPheValAlaIleArgValGlyGlyGlnLeu 499

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||:::   ||||||
Db        500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleLysLeuPro 519

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GlyAsnTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp 619
```

```
Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIle 639

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 LysLysArgLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              |||:::||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        740 GlnAsnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 759

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||            |||||||||||||||||||:::|||   ||||||||||||
Db        760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuAspIleGluGlnHisArgAla 779

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        780 LysIleGluGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||:::|||||||||||||:::|||||||||||||||||||||||||||||
Db        820 ValGlnProIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 839

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys 859

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 899

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              |||||||||||||:::||||||:::|||||||||||||   |||||||||||||||||||
Db        900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGln 919

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              |||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db        920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysValArgThrAla 939

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||:::||||||||||||||||||||||||||||||||   ||||||||||||
Db        940 HisThrAsnAspIleLysGlnLeuThrGluAlaValGlnLysIleAlaArgGluSerIle 959

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 IleIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluAla 979

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              |||||||||||||||||||||||||||||||||||||||||||   |||   ||||||||
Db       1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleProGlyValGluThrPhe 1019

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              |||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db       1020 TyrValAspGlyAlaAlaAsnArgGluThrLysMetGlyLysAlaGlyTyrValThrAsp 1039

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||||||:::|||||||||||||||||||||||||||   |||||||||||
Db       1040 ArgGlyArgGlnLysIleIleSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db       1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnValValThrAspSerGln 1079

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnLys 1099

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db       1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              |||||||||||||||||||||:::|||||||||||||||   ||||||||||||||||||
Db       1120 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

Qy       2341 TTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||
Db       1140 PheLeuAspGlyIleAsp 1145

RESULT 15
Q9WF79
ID   Q9WF79     PRELIMINARY;      PRT;  1425 AA.
AC   Q9WF79;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gag-pol polyprotein.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214383; PubMed=10196340;
RA   Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA   Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA   Essex M.;
RT   "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT   virus type 1 subtype C: a set of 23 full-length clones from
RT   Botswana.";
RL   J. Virol. 73:4427-4432(1999).
CC   -!- SIMILARITY: Belongs to peptidase family A2.
DR   EMBL; AF110969; AAD17087.1; -.
DR   HSSP; P24736; 1NCP.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008907; F:integrase activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004523; F:ribonuclease H activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015074; P:DNA integration; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
```

```
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   GO; GO:0016032; P:viral life cycle; IEA.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR001037; Integrase_C.
DR   InterPro; IPR003308; Integrase_Zn_N.
DR   InterPro; IPR001995; Peptidase_A2.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR010999; Retrovir_matrix.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR010659; RVT_connect.
DR   InterPro; IPR010661; RVT_thumb.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF00075; RnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; RVP; 1.
DR   Pfam; PF00078; RVT_1; 1.
DR   Pfam; PF06815; RVT_connect; 1.
DR   Pfam; PF06817; RVT_thumb; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW   Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   1425 AA;  160569 MW;  6DC8A1AEF3C88DA7 CRC64;

Alignment Scores:
Pred. No.:              7.95e-146      Length:        1425
Score:                  3862.00        Matches:       723
Percent Similarity:     96.31%         Conservative:  34
Best Local Similarity:  91.98%         Mismatches:    23
Query Match:            84.79%         Indels:        6
DB:                     2              Gaps:          3

US-09-610-313B-32 (1-2457) x Q9WF79 (1-1425)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTC 72
              :::||||||||||||||||||||    ::::::||||||:::||||||::::::||||||
Db        360 LeuAlaGluAlaMetSerGlnAlaAsnAsnSerAsnIleMetMetGlnLysAsnAsnPhe 379

Qy         73 AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGC 132
              |||||||||   |||||:::||||||||||||||||||||||||||||||||||||||||
Db        380 LysGlyProThrArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

Qy        133 AACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATG 192
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 AsnCysLysAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

Qy        193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGC 252
              ||||||||||||||||||:::||||||||||||||||||:::||||||||||||||||||
Db        420 LysAspCysThrGluArgArgAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439

Qy        253 AAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTG 312
              :::||||||||||||||||||||||||         |||||||||||||||||||:::|||
Db        440 GluAlaArgGluPheProSerGluGlnThrGlyAlaAsnSerProThrSerArgLysLeu 459

Qy        313 CAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAAC 372
                    ||||||:::||||||   ||||||||||   |||||||||||||||||||
Db        460 ------ArgGlyAsnAsnProCysSerGluAlaGlyAspGluArgGlnGlyThrLeuAsn 477

Qy        373 TTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATC 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        478 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 497

Qy        433 AAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCC 492
              ||||||||||||||||||||||||||||||||||||||||||||||||::::::||||||
Db        498 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 517

Qy        493 GGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTAC 552
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 GlyLysTrpArgProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 537

Qy        553 GACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCC 612
              |||||||||   ||||||||||||||||||||||||||||||||||||:::|||||||||
Db        538 AspGlnIleProIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 557

Qy        613 ACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTC 672
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        558 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPhe 577

Qy        673 CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 597

Qy        733 GTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 617

Qy        793 ATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTG 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 637

Qy        853 TTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 657

Qy        913 AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 677

Qy        973 AAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 697

Qy       1033 GACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCC 1092
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 717

Qy       1093 GGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 737

Qy       1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
              |||||||||||||||||||||||||||||||||||:::|||||||||:::||||||||||
Db        738 GlnSerSerMetThrLysIleLeuGluProPheArgAlaLysAsnProGluLeuValIle 757

Qy       1213 TACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||      ||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 777

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 ArgIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 797

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||            ||||||||||||||||||||||||
Db        798 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 817

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||   |||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        818 ValGlnSerIleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 837

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
```

```
                |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        838 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyValLysValArgGlnLeuCys 857

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        858 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProProThrGluGluAlaGlu 877

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        878 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp 897

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              |||||||||||||||:::|||||||||||||||||||||   ||||||||||||||||||
Db        898 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlyAspGlnTrpThrTyrGln 917

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 937

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        938 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleSerMetGluSerIle 957

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              |||||||||||||||   |||||||||||||||||||||||||||||||||   ||||||
Db        958 ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluAlaTrpGluAla 977

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 997

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||   ||||||:::   |||||||||||||||
Db        998 ProLeuValLysLeuTrpTyrGlnLeuGluThrGluProMetAlaGlyAlaGluThrPhe 1017

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1037

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              :::|||||||||::::::|||:::|||||||||||||||||||||      |||||||||
Db       1038 LysGlyArgGlnGluValValThrLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1057

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db       1058 AlaIleGlnLeuAlaLeuGlnAspSerGlyProGluValAsnIleValThrAspSerGln 1077

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1097

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1117

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||::::::||||||||||||      |||||||||:::||||||
Db       1118 GlyIleGlyGlyAsnGluLysValAspLysLeuValSerSerGlyIleArgGluValLeu 1137

Qy       2341 TTCCTGGACGGCATCGAT 2358
              ||||||||||||||||||
Db       1138 PheLeuAspGlyIleAsp 1143
```

Search completed: June 2, 2005, 05:05:45
Job time : 305.158 secs

GenCore version 5.1.6


```
OM nucleic - protein search, using frame_plus_n2p model

Run on:         June  2, 2005, 04:23:54 ; Search time 152.794 Seconds
                                          (without alignments)
                                          11117.427 Million cell updates/sec

Title:          US-09-610-313B-32
Perfect score:  4555
Sequence:       1 gtcgacgccaccatggccga..........gggctagcaccggtgaattc 2457

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1465611 seqs, 345679903 residues

Total number of hits satisfying chosen parameters:      2931222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09610313/runat_31052005_155138_15193/app_query.fasta_1.7893
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09610313_@CGN_1_1_550_@runat_31052005_155138_15193
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_AA:*
                1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
                17:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
                19:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                20:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description
1	3527	77.4	998	16	US-10-332-413-4	Sequence 4, Appli
2	3480.5	76.4	850	16	US-10-332-413-10	Sequence 10, Appl
3	3454	75.8	1003	14	US-10-283-847-17	Sequence 17, Appl
4	3432	75.3	1015	17	US-10-634-165-9	Sequence 9, Appli
5	3409	74.8	739	16	US-10-093-953A-26	Sequence 26, Appl
6	3409	74.8	739	16	US-10-093-953A-31	Sequence 31, Appl
7	3387	74.4	1003	16	US-10-325-468-23	Sequence 23, Appl
8	3385	74.3	1003	16	US-10-325-468-35	Sequence 35, Appl
9	3384	74.3	1003	16	US-10-325-468-9	Sequence 9, Appli
10	3384	74.3	1003	16	US-10-325-468-20	Sequence 20, Appl
11	3376	74.1	995	15	US-10-296-734-2	Sequence 2, Appli
12	3362	73.8	995	15	US-10-296-734-1470	Sequence 1470, Ap
13	3208.5	70.4	1006	15	US-10-296-734-1471	Sequence 1471, Ap
14	3140.5	68.9	1014	14	US-10-301-661A-6	Sequence 6, Appli
15	3076.5	67.5	1350	10	US-09-952-060-35	Sequence 35, Appl
16	3076.5	67.5	1350	16	US-10-380-641-35	Sequence 35, Appl
17	3076.5	67.5	1350	17	US-10-636-730-35	Sequence 35, Appl
18	2779.5	61.0	875	10	US-09-952-060-6	Sequence 6, Appli
19	2779.5	61.0	875	15	US-10-168-217A-6	Sequence 6, Appli
20	2779.5	61.0	875	16	US-10-380-641-6	Sequence 6, Appli
21	2779.5	61.0	875	17	US-10-636-730-6	Sequence 6, Appli
22	2774	60.9	850	10	US-09-952-060-2	Sequence 2, Appli
23	2774	60.9	850	15	US-10-168-217A-2	Sequence 2, Appli
24	2774	60.9	850	16	US-10-380-641-2	Sequence 2, Appli
25	2774	60.9	850	17	US-10-636-730-2	Sequence 2, Appli
26	2755.5	60.5	875	10	US-09-952-060-8	Sequence 8, Appli
27	2755.5	60.5	875	15	US-10-168-217A-8	Sequence 8, Appli
28	2755.5	60.5	875	16	US-10-380-641-8	Sequence 8, Appli
29	2755.5	60.5	875	17	US-10-636-730-8	Sequence 8, Appli
30	2750	60.4	850	10	US-09-952-060-4	Sequence 4, Appli
31	2750	60.4	850	15	US-10-168-217A-4	Sequence 4, Appli
32	2750	60.4	850	16	US-10-380-641-4	Sequence 4, Appli
33	2750	60.4	850	17	US-10-636-730-4	Sequence 4, Appli
34	2746	60.3	562	9	US-09-735-487-14	Sequence 14, Appl
35	2746	60.3	562	14	US-10-342-188-14	Sequence 14, Appl
36	2742	60.2	560	15	US-10-399-920-2	Sequence 2, Appli
37	2740	60.2	561	14	US-10-059-271-87	Sequence 87, Appl
38	2734	60.0	562	14	US-10-102-622-10	Sequence 10, Appl
39	2715	59.6	560	14	US-10-205-641-1	Sequence 1, Appli
40	2705.5	59.4	561	14	US-10-102-622-12	Sequence 12, Appl
41	2693	59.1	560	11	US-09-725-652-1	Sequence 1, Appli
42	2659	58.4	546	8	US-08-808-031A-29	Sequence 29, Appl
43	2646	58.1	979	14	US-10-271-181B-118	Sequence 118, App
44	2553	56.0	851	14	US-10-059-271-88	Sequence 88, Appl
45	2528	55.5	999	15	US-10-346-000A-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-10-332-413-4
; Sequence 4, Application US/10332413
; Publication No. US20040116660A1
; GENERAL INFORMATION:
;  APPLICANT: Johnstone, Robert Edward
;  APPLICANT:  Swanstrom, Ronald Ivar
;  APPLICANT:  Morris, Lynn
;  APPLICANT:  Karim, Salim Abdool
;  APPLICANT:  Williamson, Carolyn
;  TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H
;  TITLE OF INVENTION:  Subtype Isolates, Their Genes nd Modifications and Derivatives T
;  FILE REFERENCE: 45669-281993
;  CURRENT APPLICATION NUMBER: US/10/332,413
;  CURRENT FILING DATE:  2003-09-12
;  PRIOR APPLICATION NUMBER: PCT/IB01/01208
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/216,995
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: ZA 2000/3437
;  PRIOR FILING DATE: 2000-07-10
;  PRIOR APPLICATION NUMBER: ZA 2000/4924
```

```
;  PRIOR FILING DATE: 2000-09-15
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 998
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-10-332-413-4

Alignment Scores:
Pred. No.:              1.14e-163     Length:        998
Score:                  3527.00       Matches:       662
Percent Similarity:     96.51%        Conservative:  29
Best Local Similarity:  92.46%        Mismatches:    21
Query Match:            77.43%        Indels:        4
DB:                     16            Gaps:          2

US-09-610-313B-32 (1-2457) x US-10-332-413-4 (1-998)

Qy        223 TTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAAC 282
Db          1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20

Qy        283 CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAG 342
Db         21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu 40

Qy        343 GCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCC 402
Db         41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro 60

Qy        403 CTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGAC 462
Db         61 LeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp 80

Qy        463 GACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGC 522
Db         81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 100

Qy        523 ATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAG 582
Db        101 IleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLys 120

Qy        583 AAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATG 642
Db        121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140

Qy        643 CTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTG 702
Db        141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160

Qy        703 AAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAG 762
Db        161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180

Qy        763 ATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATC 822
Db        181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200

Qy        823 GGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAG 882
Db        201 GlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220

Qy        883 TGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG 942
Db        221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240

Qy        943 CAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGAC 1002
Db        241 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp 260

Qy       1003 GTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTC 1062
Db        261 ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 280

Qy       1063 ACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCC 1122
Db        281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 300

Qy       1123 CAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCC 1182
Db        301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320

Qy       1183 TTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGC 1236
Db        321 PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 340

Qy       1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
Db        341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeu 360

Qy       1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------ 1350
Db        361 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380

Qy       1351 CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1410
Db        381 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400

Qy       1411 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1470
Db        401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420

Qy       1471 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1530
Db        421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440

Qy       1531 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1590
Db        441 AspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu 460

Qy       1591 CGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAG 1650
Db        461 LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln 480

Qy       1651 AAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAG 1710
Db        481 LysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 500

Qy       1711 ACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAG 1770
Db        501 ThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThrGlu 520

Qy       1771 GCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGC 1830
Db        521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540

Qy       1831 CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGG 1890
Db        541 LeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThrTrp 560

Qy       1891 ATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1950
Db        561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 580

Qy       1951 AAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACC 2010
Db        581 LysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 600

Qy       2011 AAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACC 2070
Db        601 LysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeuSer 620

Qy       2071 GAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGC 2130
Db        621 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640
```

```
Qy       2131 AGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCC 2190
              ||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        641 SerGluValAsnIleIleThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 660

Qy       2191 GACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG 2250
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        661 AspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluArg 680

Qy       2251 GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG 2310
              |||||||||||||||||||||||||     |||:::   |||||||||    :::||||||
Db        681 ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700

Qy       2311 CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||||||||:::|||||||||||||||||||    |||||||||||||||
Db        701 LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716
```

RESULT 2
US-10-332-413-10
; Sequence 10, Application US/10332413
; Publication No. US20040116660A1
; GENERAL INFORMATION:
; APPLICANT: Johnstone, Robert Edward
; APPLICANT: Swanstrom, Ronald Ivar
; APPLICANT: Morris, Lynn
; APPLICANT: Karim, Salim Abdool
; APPLICANT: Williamson, Carolyn
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H
; TITLE OF INVENTION: Subtype Isolates, Their Genes nd Modifications and Derivatives T
; FILE REFERENCE: 45669-281993
; CURRENT APPLICATION NUMBER: US/10/332,413
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: PCT/IB01/01208
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/216,995
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: ZA 2000/3437
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: ZA 2000/4924
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 850
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-332-413-10

Alignment Scores:

Pred. No.:	2.06e-161	Length:	850
Score:	3480.50	Matches:	664
Percent Similarity:	90.53%	Conservative:	34
Best Local Similarity:	86.12%	Mismatches:	46
Query Match:	76.41%	Indels:	27
DB:	16	Gaps:	6

US-09-610-313B-32 (1-2457) x US-10-332-413-10 (1-850)

```
Qy         97 TGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAG 156
                |||         ||||||                              |||        |||
Db         84 CysAlaThrValGlyLys------------GlyAspArgCysGlyProLeuArgTyrTyr 99

Qy        157 GGCTGCTGGAAG-----------------------TGCGGCAAGGAGGGCCACCAGATG 192
                       |||:::                        |||::::::|||
Db        100 AlaSerTrpArgLysGlyAspValLeuGlnGlyAspValGlyArgGlnGlyPheProSer 119

Qy        193 AAGGACTGCACCGAGCGCCAG------------GCCAACTTCTTCCGCGAGGACCTG--- 237
                    |||       :::|||:::                 :::        |||       |||
Db        120 HisAspValValLysArgArgProValProSerLeuHisAlaCysArgSerThrLeuGlu 139

Qy        238 ------------------GCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
                                :::||||||||||||    ||||||:::||||||||||||||
Db        140 AspProArgValProSerSerPheProGlnGlyProAlaArgGlnPheProSerGluGln 159

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
                |||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        160 ThrArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAspAsnProArgSer 179

Qy        340 GAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGC 399
              |||    ||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db        180 GluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg 199

Qy        400 CCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCC 459
              ||||||||||||||||:::|||||||||    :::|||||||||||||||||||||||||
Db        200 ProLeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAla 219

Qy        460 GACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGC 519
              |||||||||||||||||||||::::::::::||||||||||||||||||||||||||||||
Db        220 AspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGly 239

Qy        520 GGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGC 579
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGly 259

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
              |||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        260 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 279

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 299

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluVal 319

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 LysIleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLys 339

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
              |||||||||||||||||||||||||||:::||||||||||||||:::|||||||||
Db        340 IleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysGluAspSerThr 359

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 379

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 399

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
              |||||||||||||||||||||||||||||||||||||||    |||||||||||||||||
Db        400 AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAla 419

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 439

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
              ||||||||||||||||||||||:::|||||||||:::|||||||||||||||||||||
Db        440 ProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSerMetThrLysIleLeuGlu 459

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTG 1233
              |||||||||||:::||||||||||||||||||||||||         |||    |||||||||
Db        460 ProPheArgAlaLysAsnProGluIleValIleTyrGlnTyrMetAlaAlaLeuTyrVal 479

Qy       1234 GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTG 1293
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        480 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeu 499

Qy       1294 CTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG--- 1350
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 519
```

```
Qy       1351 ---CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAG 1407
                       ||||||||||||||||||||||||||||||||||||:::|||||||||||
Db        520 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLys 539

Qy       1408 GAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG 1467
              :::||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        540 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpThrSerGln 559

Qy       1468 ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTG 1527
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        560 IleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 579

Qy       1528 ACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATC 1587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIle 599

Qy       1588 CTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATC 1647
              |||:::||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        600 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 619

Qy       1648 CAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTG 1707
              |||||||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        620 GlnLysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 639

Qy       1708 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACC 1767
              ||||||||||||||||||||   ||||||   ||||||||||||||||||||||||||||
Db        640 LysThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThr 659

Qy       1768 GAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTC 1827
              ||||||||||||||||::::::||||||||||||   |||||||||||||||||||||||
Db        660 GluAlaValGlnLysIleSerLeuGluSerIleValThrTrpGlyLysThrProLysPhe 679

Qy       1828 CGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACC 1887
              ||||||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db        680 ArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThr 699

Qy       1888 TGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 TrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeu 719

Qy       1948 GAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAG 2007
              ||||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db        720 GluLysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 739

Qy       2008 ACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTG 2067
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        740 ThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeu 759

Qy       2068 ACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGC 2127
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 SerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 779

Qy       2128 GGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAG 2187
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 GluSerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln 799

Qy       2188 CCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAG 2247
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 ProAspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlu 819

Qy       2248 AAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGAC 2307
              :::   ||||||||||||||||||||||||||||||||||||:::||||||:::|||
Db        820 ArgAlaTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAspGluGlnValAsp 839

Qy       2308 AAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||   ||||||||||||||||||
Db        840 LysLeuValSerSerGlyIleArgLysValLeu 850
```

RESULT 3
US-10-283-847-17
; Sequence 17, Application US/10283847
; Publication No. US20030162720A1
; GENERAL INFORMATION:
; APPLICANT: Nicolette, Charles A.
; APPLICANT: Walker, Bruce
; TITLE OF INVENTION: THERAPEUTIC ANTI-HIV (IV9) COMPOUNDS
; FILE REFERENCE: GZ 2112.00
; CURRENT APPLICATION NUMBER: US/10/283,847
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 60/345,116
; PRIOR FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus
US-10-283-847-17

Alignment Scores:

Pred. No.:	4.06e-160	Length:	1003
Score:	3454.00	Matches:	651
Percent Similarity:	95.01%	Conservative:	34
Best Local Similarity:	90.29%	Mismatches:	28
Query Match:	75.83%	Indels:	8
DB:	14	Gaps:	4

US-09-610-313B-32 (1-2457) x US-10-283-847-17 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||   |||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                ||||||||||||||||   ||||||||||||||         |||   :::|||
Db         21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 |||||||||||||:::||||||||||||:::      ||||||||||||:::|||
Db         41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              |||||||||||||||||||:::||||||:::|||||||||:::|||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              ||||||||||||   |||||||||||||||||||:::|||||||||||||||||||||||
Db        121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              ||||||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              ||||||||||||||||||||||||||        ||||||   |||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
```

```
Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db        321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db        361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db        421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

Qy       1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460

Qy       1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db        481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

Qy       1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520

Qy       1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db        521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540

Qy       1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db        541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620

Qy       2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db        621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640

Qy       2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
Db        661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680

Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720

Qy       2356 GAT 2358
Db        721 Asp 721
```

RESULT 4
US-10-634-165-9
; Sequence 9, Application US/10634165
; Publication No. US20050095581A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Nancy T
; APPLICANT: Gallo, Robert C.
; APPLICANT: Wong-Staal, Flossie
; TITLE OF INVENTION: DETECTION OF HIV-1 DNA
; FILE REFERENCE: 223695
; CURRENT APPLICATION NUMBER: US/10/634,165
; CURRENT FILING DATE: 2003-08-05
; PRIOR APPLICATION NUMBER: 08/463,028
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 06/693,866
; PRIOR FILING DATE: 1985-01-23
; PRIOR APPLICATION NUMBER: 06/659,339
; PRIOR FILING DATE: 1984-10-10
; PRIOR APPLICATION NUMBER: 06/643,306
; PRIOR FILING DATE: 1984-08-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
;   LENGTH: 1015
;   TYPE: PRT
;   ORGANISM: T cell leukemia-lymphoma virus (HTLV)
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1015)
;   OTHER INFORMATION: /note= "pol protein of HTLV-III"
US-10-634-165-9

Alignment Scores:

Pred. No.:	4.78e-159	Length:	1015
Score:	3432.00	Matches:	650
Percent Similarity:	93.18%	Conservative:	33
Best Local Similarity:	88.68%	Mismatches:	30
Query Match:	75.35%	Indels:	20
DB:	17	Gaps:	5

US-09-610-313B-32 (1-2457) x US-10-634-165-9 (1-1015)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTC------------ 267
Db          1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
```

```
Qy      268 -----------------------CCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGC 303
Db       21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40

Qy      304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db       41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy      358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db       61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy      412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db       81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy      472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db      101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy      532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db      121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy      592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db      141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy      652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db      161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy      712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db      181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy      772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db      201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy      832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db      221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy      892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db      241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy      952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db      261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy     1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db      281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy     1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db      301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy     1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db      321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy     1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db      341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

Qy     1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db      361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy     1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAG 1359
Db      381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy     1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
Db      401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy     1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
Db      421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440

Qy     1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
Db      441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy     1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
Db      461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy     1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
Db      481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy     1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
Db      501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy     1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
Db      521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540

Qy     1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
Db      541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy     1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
Db      561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy     1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
Db      581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy     1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
Db      601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy     2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
Db      621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy     2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
Db      641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy     2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
Db      661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy     2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
Db      681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy     2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
Db      701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy     2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db      721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 5

```
US-10-093-953A-26
; Sequence 26, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
;  APPLICANT: Robinson, Harriet L.
;  APPLICANT:  Smith, James M.
;  APPLICANT:  Hua, Jian
;  APPLICANT:  Moss, Bernard
;  APPLICANT:  Amara, Rama
;  APPLICANT:  Wyatt, Linda
;  APPLICANT:  Earl, Patricia
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
;  TITLE OF INVENTION:  AN IMMUNE RESPONSE
;  FILE REFERENCE: 12804-005002
;  CURRENT APPLICATION NUMBER: US/10/093,953A
;  CURRENT FILING DATE:  2002-03-08
;  PRIOR APPLICATION NUMBER: US 60/186,364
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 60/251,083
;  PRIOR FILING DATE: 2000-12-01
;  PRIOR APPLICATION NUMBER: US 09/798,675
;  PRIOR FILING DATE: 2001-03-02
;  PRIOR APPLICATION NUMBER: PCT/US01/06795
;  PRIOR FILING DATE: 2001-03-02
;  PRIOR APPLICATION NUMBER: US 60/324,845
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/325,004
;  PRIOR FILING DATE: 2001-09-26
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
;   LENGTH: 739
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: protein encoded by construct of vaccine vector
;   OTHER INFORMATION: pGA2 and insert JS2 expressing clade HIV-1 VL
US-10-093-953A-26

Alignment Scores:
Pred. No.:              6.15e-158     Length:         739
Score:                  3409.00       Matches:        647
Percent Similarity:     92.91%        Conservative:   34
Best Local Similarity:  88.27%        Mismatches:     32
Query Match:            74.84%        Indels:         20
DB:                     16            Gaps:           5

US-09-610-313B-32 (1-2457) x US-10-093-953A-26 (1-739)

Qy     220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db       1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy     280 AACCGC------------------------------------GCCAACAGCCCCACCAGC 303
Db      21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40

Qy     304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
Db      41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy     358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
Db      61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy     412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
Db      81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy     472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
Db     101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy     532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
Db     121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy     592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db     141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy     652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
Db     161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy     712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
Db     181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy     772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
Db     201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

Qy     832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
Db     221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy     892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db     241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy     952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
Db     261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy    1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
Db     281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy    1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
Db     301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy    1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db     321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy    1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
Db     341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360

Qy    1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
Db     361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy    1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAG 1359
Db     381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy    1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
Db     401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy    1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
Db     421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440

Qy    1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
Db     441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy    1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
Db     461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy    1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
```

```
                  ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db         481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy        1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy        1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
               ||||||:::||||||   |||||||||||||||||||||   ||||||||||||||||||
Db         521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540

Qy        1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
               ||||||      |||||||||||||||||||||||||||||||||||:::||||||||||
Db         541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy        1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
               ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db         561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy        1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy        1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
               |||:::||||||||||||||||||||||||||||||||||||||||||||||:::|||||
Db         601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy        2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
               ||||||||||||||||||:::::::||||||||||||:::|||   ||||||   ||||||
Db         621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy        2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
               ||||||||||||:::|||||||||||||   |||||||||||||||||||||   ||||||
Db         641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy        2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy        2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy        2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
               :::||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db         701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy        2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
                  ||||||||||:::||||||||||||||||||||||
Db         721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 6
US-10-093-953A-31
; Sequence 31, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; APPLICANT: Moss, Bernard
; APPLICANT: Amara, Rama
; APPLICANT: Wyatt, Linda
; APPLICANT: Earl, Patricia
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; TITLE OF INVENTION: AN IMMUNE RESPONSE
; FILE REFERENCE: 12804-005002
; CURRENT APPLICATION NUMBER: US/10/093,953A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/798,675
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/325,004
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 739
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: protein encoded by construct of vaccine vector
;   OTHER INFORMATION: pGA1 and vaccine insert expressing clade B HIV-1
;   OTHER INFORMATION: Gag-Pol
US-10-093-953A-31

```
Alignment Scores:
Pred. No.:              6.15e-158     Length:         739
Score:                  3409.00       Matches:        647
Percent Similarity:     92.91%        Conservative:   34
Best Local Similarity:  88.27%        Mismatches:     32
Query Match:            74.84%        Indels:         20
DB:                     16            Gaps:           5

US-09-610-313B-32 (1-2457) x US-10-093-953A-31 (1-739)

Qy         220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
               ||||||||||||||||||||||||   |||||||||||||||||||||   |||||||||
Db           1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy         280 AACCGC------------------------------------GCCAACAGCCCCACCAGC 303
                 |||                                     |||||||||||||||
Db          21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40

Qy         304 CGCGAGCTGCAGGTG------CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGC 357
               |||||||||||||||      |||   :::|||   ||||||||||||||:::|||
Db          41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

Qy         358 CAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGC 411
               |||||||||:::      ||||||||||||||||||||||||||||||||||||||:::
Db          61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

Qy         412 ATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTG 471
               ||||||:::|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db          81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100

Qy         472 CTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGC 531
               |||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db         101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120

Qy         532 TTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATC 591
               ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db         121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140

Qy         592 GGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
               ||||||||||||:::|||||||||||||||||||||||||||||||||:::|||||||||
Db         141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

Qy         652 CTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAG 711
               :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

Qy         712 CCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCC 771
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200

Qy         772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAG 831
               |||      ||||||||||||||||||||||||||||||:::||||||||||||||||||
Db         201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
```

```
Qy        832 AACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

Qy        892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

Qy        952 ATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGAC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280

Qy       1012 GCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

Qy       1072 AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGG 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

Qy       1132 AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||:::
Db        321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

Qy       1192 CGCAACCCCGAGATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTG 1245
              :::||||||:::|||||||||||||||      ||||||||||||||||||||||||
Db        341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360

Qy       1246 GAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGC 1305
              |||||||||||||||||   ||||||||||||||||||:::|||||||||||||||||||
Db        361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

Qy       1306 TTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAG 1359
                 ||||||||||||||||||||||||||||||||||||||||||         |||
Db        381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

Qy       1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACC 1419
              |||||||||||||||||||||||||||||||||||   |||||||||:::||||||||||
Db        401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

Qy       1420 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGC 1479
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440

Qy       1480 ATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTG 1539
              ||||||||||||||||||||||||||||||||||||   ||||||||||||:::::::::
Db        441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

Qy       1540 CCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1599
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

Qy       1600 GTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGC 1659
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500

Qy       1660 CACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1719
                       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520

Qy       1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAG 1779
              ||||||:::||||||   |||||||||||||||||||||   ||||||||||||||||||
Db        521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540

Qy       1780 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATC 1839
              ||||||      |||||||||||||||||||||||||||||||||:::||||||||||||
Db        541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560

Qy       1840 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1899
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db        561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580

Qy       1900 TGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600

Qy       1960 ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGC 2019
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620

Qy       2020 AAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACC 2079
              ||||||||||||||||||::::::||||||||||||:::|||   ||||||   ||||||
Db        621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640

Qy       2080 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTG 2139
              ||||||||||||:::||||||||||||   ||||||||||||||||||||   |||||||
Db        641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy       2140 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGC 2199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680

Qy       2200 GAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTG 2259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700

Qy       2260 AGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2319
              :::|||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720

Qy       2320 AAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
                 |||||||||:::||||||||||||||||||||||
Db        721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 7
US-10-325-468-23
; Sequence 23, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone 1.10 protein Pol
US-10-325-468-23

```
Alignment Scores:
Pred. No.:               7.38e-157      Length:         1003
Score:                   3387.00        Matches:        634
Percent Similarity:      94.17%         Conservative:   45
Best Local Similarity:   87.93%         Mismatches:     34
Query Match:             74.36%         Indels:         8
DB:                      16             Gaps:           4
```

US-09-610-313B-32 (1-2457) x US-10-325-468-23 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              |||||||||||||||||||||||||||||||||||||||:::|||   ||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                 ||||||||||||||      ||||||   ||||||      |||   :::|||
Db         21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
```

```
Qy    334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db     41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy    388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db     61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy    448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db     81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy    508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db    101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy    568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db    121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy    628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db    141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy    688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db    161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy    748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db    181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy    808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db    201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy    868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db    221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy    928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db    241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy    988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db    261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy   1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db    281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy   1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db    301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy   1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db    321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy   1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db    341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy   1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db    361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy   1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
Db    381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy   1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db    401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy   1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db    421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy   1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db    441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy   1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db    461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy   1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db    481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy   1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db    501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgSerAlaHisThrAsnAspVal 520

Qy   1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db    521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy   1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db    541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy   1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db    561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy   1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db    581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy   1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db    601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy   2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db    621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy   2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
Db    641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy   2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
Db    661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy   2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
Db    681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy   2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
Db    701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy   2356 GAT 2358
Db    721 Glu 721
```

RESULT 8
US-10-325-468-35
; Sequence 35, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:

```
;  APPLICANT: Soong, Nay Wei
;  APPLICANT:  Pekrun, Katja
;  APPLICANT:  Shibata, Riri
;  TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
;  TITLE OF INVENTION:  MODELS OF HIV-1 PATHOGENESIS
;  FILE REFERENCE: 0166.210US
;  CURRENT APPLICATION NUMBER: US/10/325,468
;  CURRENT FILING DATE:  2002-12-19
;  PRIOR APPLICATION NUMBER: US 60/343,524
;  PRIOR FILING DATE: 2001-12-21
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone P8A26 protein Pol
US-10-325-468-35

Alignment Scores:
Pred. No.:              9.23e-157      Length:         1003
Score:                  3385.00        Matches:        635
Percent Similarity:     94.17%         Conservative:   44
Best Local Similarity:  88.07%         Mismatches:     34
Query Match:            74.31%         Indels:         8
DB:                     16             Gaps:           5

US-09-610-313B-32 (1-2457) x US-10-325-468-35 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG---CGCGGCGAC---AACCCC 333
Db         21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgGlyAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db         81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
Db        321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
Db        361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy       1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db        481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy       1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520

Qy       1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
```

```
Qy       1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
              ||||||||||||||||||:::||||||||||||||||||||||||   |||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy       2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
              :::||||||||||||:::|||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
              ||||||||||||||||||||||||:::|||:::|||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuThrTrpIleProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
              ||||||:::|||||||||||||||   |||||||||:::|||||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy       2356 GAT 2358
              :::
Db        721 Glu 721
```

RESULT 9
US-10-325-468-9
; Sequence 9, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone 1.4 protein Pol;
;   OTHER INFORMATION: clone 1.26 protein Pol; clone P10.21 protein Pol;
;   OTHER INFORMATION: clone P10.26 protein Pol
US-10-325-468-9

```
Alignment Scores:
Pred. No.:              1.03e-156     Length:         1003
Score:                  3384.00       Matches:        634
Percent Similarity:     94.04%        Conservative:   44
Best Local Similarity:  87.93%        Mismatches:     35
Query Match:            74.29%        Indels:         8
DB:                     16            Gaps:           4
```

US-09-610-313B-32 (1-2457) x US-10-325-468-9 (1-1003)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
                 ||||||||||||||      ||||||    ||||||       |||    :::|||
Db         21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
                 ||||||||||||||:::||||||||||||:::      ||||||||||||||||||
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
              ||||||||||||||||||:::||||||:::|||||||||:::||||||||||||||||||
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
              ||||||||||||||||||||||||||||:::|||      ||||||:::|||||||||
Db         81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
              :::|||||||||    ||||||:::||||||||||||:::||||||||||||||||||||
Db        121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
              |||||||||:::|||||||||:::||||||||||||||||||||||||||||||||||||
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
              |||||||||||||||||||||||||||       ||||||    ||||||||||||||||||
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::|||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              |||    ||||||||||||||    :::||||||:::|||:::|||||||||
Db        321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                 ||||||||||||||||||||||||||||||||||||||     ||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
```

```
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy       1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
Db        481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy       1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520

Qy       1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy       2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy       2356 GAT 2358
              :::
Db        721 Glu 721
```

RESULT 10
US-10-325-468-20

```
; Sequence 20, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
;  APPLICANT: Soong, Nay Wei
;  APPLICANT:  Pekrun, Katja
;  APPLICANT:  Shibata, Riri
;  TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
;  TITLE OF INVENTION:  MODELS OF HIV-1 PATHOGENESIS
;  FILE REFERENCE: 0166.210US
;  CURRENT APPLICATION NUMBER: US/10/325,468
;  CURRENT FILING DATE:  2002-12-19
;  PRIOR APPLICATION NUMBER: US 60/343,524
;  PRIOR FILING DATE: 2001-12-21
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: recombinant / chimeric sequence: clone 1.27 protein Pol
US-10-325-468-20

Alignment Scores:
Pred. No.:              1.03e-156        Length:          1003
Score:                  3384.00          Matches:         634
Percent Similarity:     94.04%           Conservative:    44
Best Local Similarity:  87.93%           Mismatches:      35
Query Match:            74.29%           Indels:          8
DB:                     16               Gaps:            4

US-09-610-313B-32 (1-2457) x US-10-325-468-20 (1-1003)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG------CGCGGCGACAACCCC 333
Db         21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACC 387
Db         41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy        388 CTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTG 447
Db         61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy        448 GACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCC 507
Db         81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100

Qy        508 AAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATC 567
Db        101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120

Qy        568 GAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATC 627
Db        121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140

Qy        628 ATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATC 687
Db        141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160

Qy        688 GAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCC 747
Db        161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180

Qy        748 CTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGC 807
Db        181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
```

```
Qy        808 AAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAG 867
              |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

Qy        868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||::::::||||||
Db        221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240

Qy        928 GACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSer 260

Qy        988 GTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGC 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280

Qy       1048 AAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300

Qy       1108 TACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACC 1167
              |||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

Qy       1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCC 1221
              |||   ||||||||||||||      :::||||||:::|||:::||||||||
Db        321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340

Qy       1222 CCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTG 1281
                    ||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360

Qy       1282 CGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCC 1341
              |||:::|||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380

Qy       1342 CCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1395
              |||||||||            |||||||||||||||||||||||||||||||||||||
Db        381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

Qy       1396 CTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAC 1455
              ||||||||||::::||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420

Qy       1456 TGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGC 1515
              ||||||||||||||||||   ||||||||||||:::||||||||||||||||||||||||
Db        421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440

Qy       1516 GCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG 1575
                 ||||||||||||::::::::|||||||||:::|||||||||||||||||||||||||
Db        441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460

Qy       1576 AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTG 1635
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480

Qy       1636 GTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCC 1695
              :::   ||||||||||||||||||         |||||||||||||:::||||||||||
Db        481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500

Qy       1696 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTG 1755
              ||||||||||||||||||||||||||||||||||   |||   |||||||||||||||||
Db        501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520

Qy       1756 AAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1815
              ||||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db        521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540

Qy       1816 ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
                 |||||||:::||||||||||||||||||||||||||||||||||||||:::|||
Db        541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560

Qy       1876 TGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTG 1935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy       1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 1995
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy       1996 GCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAG 2055
              ||||||||||||||||:::|||||||||||||||||||||   ||||||||||||||
Db        601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620

Qy       2056 ATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
              :::|||||||||||:::||||||||||||||||||||||||||||||||||   ||||||
Db        621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640

Qy       2116 CTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
              ||||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db        641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660

Qy       2176 ATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTG 2235
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

Qy       2236 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC 2295
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy       2296 GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATC 2355
              ||||||:::|||||||||||||||||   ||||||||||:::||||||||||||||||||
Db        701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720

Qy       2356 GAT 2358
              :::
Db        721 Glu 721
```

RESULT 11
US-10-296-734-2
; Sequence 2, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 995
;   TYPE: PRT
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: POL consensus polypeptide
US-10-296-734-2

```
Alignment Scores:
Pred. No.:              2.53e-156     Length:         995
Score:                  3376.00       Matches:        636
Percent Similarity:     93.89%        Conservative:   40
Best Local Similarity:  88.33%        Mismatches:     30
Query Match:            74.12%        Indels:         14
DB:                     15            Gaps:           5

US-09-610-313B-32 (1-2457) x US-10-296-734-2 (1-995)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||||:::|||||||||   ||||||||||||||||||||   |||||||||
Db          1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
```

```
Db      21 ThrGlyAlaAsnSerSerAlaSerArgLysLeu---------GlyAspGly--------- 34

Qy     340 GAGGCCGGCGCCGAGCGCCAGGGCACC---------CTGAACTTCCCCAGATCACCCTG 390
Db      35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheProGlnIleThrLeu 53

Qy     391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
Db      54 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 73

Qy     451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
Db      74 ThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLys 93

Qy     511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
Db      94 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 113

Qy     571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
Db     114 IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133

Qy     631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
Db     134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153

Qy     691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
Db     154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 173

Qy     751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
Db     174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGluGlyLys 193

Qy     811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
Db     194 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 213

Qy     871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
Db     214 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233

Qy     931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
Db     234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 253

Qy     991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
Db     254 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLys 273

Qy    1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
Db     274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 293

Qy    1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
Db     294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313

Qy    1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
Db     314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTyrGlnTyrMetAspAsp 333

Qy    1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
Db     334 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353

Qy    1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
Db     354 AlaHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373

Qy    1345 TTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1398
Db     374 PheLeuTrpMetGlyTyrGluLeuHisProAspArgTrpThrValGlnProIleGluLeu 393

Qy    1399 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1458
Db     394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413

Qy    1459 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1518
Db     414 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 433

Qy    1519 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1578
Db     434 LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuGluGluAsn 453

Qy    1579 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1638
Db     454 ArgGluIleLeuArgGluProValHisGlyValTyrTyrAspProSerLysAspLeuVal 473

Qy    1639 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1698
Db     474 AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPhe 493

Qy    1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1758
Db     494 LysAsnLeuLysThrGlyLysTyrSerArgLysArgSerAlaHisThrAsnAspValArg 513

Qy    1759 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1818
Db     514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533

Qy    1819 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878
Db     534 ProLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTyrTrp 553

Qy    1879 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1938
Db     554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 573

Qy    1939 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 1998
Db     574 TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 593

Qy    1999 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2058
Db     594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613

Qy    2059 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118
Db     614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633

Qy    2119 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2178
Db     634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653

Qy    2179 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2238
Db     654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673

Qy    2239 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2298
Db     674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693

Qy    2299 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
Db     694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713
```

RESULT 12
US-10-296-734-1470
; Sequence 1470, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A

```
;  TITLE OF INVENTION: Synthetic molecules and uses therefor
;  FILE REFERENCE: Savine
;  CURRENT APPLICATION NUMBER: US/10/296,734
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: AU PQ7761/00
;  PRIOR FILING DATE: 2000-05-26
;  NUMBER OF SEQ ID NOS: 1507
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1470
;   LENGTH: 995
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus
US-10-296-734-1470

Alignment Scores:
Pred. No.:              1.21e-155     Length:        995
Score:                  3362.00       Matches:       634
Percent Similarity:     93.61%        Conservative:  40
Best Local Similarity:  88.06%        Mismatches:    32
Query Match:            73.81%        Indels:        14
DB:                     15            Gaps:          5

US-09-610-313B-32 (1-2457) x US-10-296-734-1470 (1-995)

Qy       220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db         1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy       280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGC 339
Db        21 ThrGlyAlaAsnSerSerAlaSerArgLysLeu---------GlyAspGly--------- 34

Qy       340 GAGGCCGGCGCCGAGCGCCAGGGCACC---------CTGAACTTCCCCCAGATCACCCTG 390
Db        35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheProGlnIleThrLeu 53

Qy       391 TGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
Db        54 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 73

Qy       451 ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAG 510
Db        74 ThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLys 93

Qy       511 ATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAG 570
Db        94 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 113

Qy       571 ATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATC 630
Db       114 IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133

Qy       631 GGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAG 690
Db       134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153

Qy       691 ACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTG 750
Db       154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 173

Qy       751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810
Db       174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGluGlyLys 193

Qy       811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870
Db       194 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 213

Qy       871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930
Db       214 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233

Qy       931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990
Db       234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal 253

Qy       991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050
Db       254 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLys 273

Qy      1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110
Db       274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 293

Qy      1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170
Db       294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313

Qy      1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG------GCCCCC 1224
Db       314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTyrGlnTyrMetAspAsp 333

Qy      1225 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1284
Db       334 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353

Qy      1285 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1344
Db       354 AlaHisLeuLeuSerTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373

Qy      1345 TTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1398
Db       374 PheLeuTrpMetGlyTyrGluLeuHisProAspArgTrpThrValGlnProIleGluLeu 393

Qy      1399 CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGG 1458
Db       394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413

Qy      1459 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1518
Db       414 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 433

Qy      1519 AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAC 1578
Db       434 LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuGluGluAsn 453

Qy      1579 CGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTG 1638
Db       454 ArgGluIleLeuArgIleProValHisGlyValTyrTyrAspProSerLysAspLeuVal 473

Qy      1639 GCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTC 1698
Db       474 AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPhe 493

Qy      1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAG 1758
Db       494 LysAsnLeuLysThrGlyLysTyrSerArgLysArgSerAlaHisThrAsnAspValArg 513

Qy      1759 CAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACC 1818
Db       514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533

Qy      1819 CCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878
Db       534 ProLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTyrTrp 553

Qy      1879 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGG 1938
Db       554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 573

Qy      1939 TACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC 1998
Db       574 TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 593

Qy      1999 AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATC 2058
```

```
Db        594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613

Qy       2059 GTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118
              :::||||||||||||||||||||||||||||||||||||    ||||||    |||||||||
Db        614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633

Qy       2119 CAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATC 2178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653

Qy       2179 CAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATC 2238
              |||||||||||||:::|||||||||||:::|||:::||||||||||||:::||||||
Db        654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673

Qy       2239 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAG 2298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693

Qy       2299 CAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||:::|||||||||||      ||||||||||||||||||||||||||||||||:::
Db        694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713
```

RESULT 13
US-10-296-734-1471
; Sequence 1471, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1471
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: POL Consensus A
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (15)..(15)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (22)..(23)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (26)..(26)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (30)..(30)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (36)..(36)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (38)..(39)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (41)..(44)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (46)..(46)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (48)..(48)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (51)..(51)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (53)..(55)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (58)..(58)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (73)..(73)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (75)..(75)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (105)..(105)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (130)..(130)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (173)..(173)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (194)..(194)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (197)..(197)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (255)..(255)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (280)..(281)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (293)..(293)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (300)..(300)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (316)..(316)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (320)..(320)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (331)..(333)

```
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (335)..(335)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (365)..(365)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (369)..(369)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (373)..(373)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (402)..(403)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (434)..(434)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (439)..(439)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (452)..(452)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (470)..(470)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (480)..(480)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (487)..(487)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (514)..(514)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (534)..(535)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (544)..(544)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (553)..(553)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (558)..(558)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (593)..(593)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (610)..(610)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (675)..(675)
;     OTHER INFORMATION: X = unknown amino acid
;     FEATURE:

Alignment Scores:
Pred. No.:              3.55e-148     Length:         1006
Score:                  3208.50       Matches:        614
Percent Similarity:     88.26%        Conservative:   25
Best Local Similarity:  84.81%        Mismatches:     74
Query Match:            70.44%        Indels:         11
DB:                     15            Gaps:           4

US-09-610-313B-32 (1-2457) x US-10-296-734-1471 (1-1006)

Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
              ||||||||||:::||||||||   ||||||:::||||||   |||   ||||||||
Db          1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArg***PheSerSerGluGln 20

Qy        280 AACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCC------ 333
                    ||||||   |||||||||   |||         |||   :::
Db         21 Thr******AsnSer***ThrSerArg***LeuTrpAspGlyGly***Asp******Leu 40

Qy        334 CGCAGCGAGGCCGGCGCCGAGCGCCAGGGC---------------ACCCTGAACTTCCCC 378
                          |||   |||   ||||||                :::      ||||||
Db         41 ************Gly***Glu***GlnGly***Gly*********SerPhe***PhePro 60

Qy        379 CAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAG 438
              |||||||||||||||||||||||||||||::::::   :::   ||||||:::   |||
Db         61 GlnIleThrLeuTrpGlnArgProLeuValThrVal***Ile***GlyGlnLeuIleGlu 80

Qy        439 GCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAG 498
              ||||||||||||||||||||||||||||||||||||:::::::::||||||||||||
Db         81 AlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLys 100

Qy        499 TGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAG 558
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        101 TrpLysProLys***IleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGln 120

Qy        559 ATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCC 618
              |||||||||||||||||||||||||||   ||||||||||||||||:::||||||||||
Db        121 IleLeuIleGluIleCysGlyLysLys***IleGlyThrValLeuValGlyProThrPro 140

Qy        619 GTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATC 678
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        141 ValAsnIleIleGlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIle 160

Qy        679 AGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAG 738
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        161 SerProIleGluThrValProValLysLeuLysPro***MetAspGlyProLysValLys 180

Qy        739 CAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAG 798
              |||||||||||||||||||||||||||||||||||||   ||||||   |||||||||
Db        181 GlnTrpProLeuThrGluGluLysIleLysAlaLeuThr***IleCys***GluMetGlu 200

Qy        799 AAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCC 858
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAla 220

Qy        859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 IleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys 240

Qy        919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAG 978
              |||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db        241 ArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis***AlaGlyLeuLysLys 260

Qy        979 AAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAG 1038
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 LysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAsp*** 280

Qy       1039 GACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATC 1098
```

```
Db        281 ***PheArgLysTyrThrAlaPheThrIleProSer***AsnAsnGluThrProGly*** 300

Qy       1099 CGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGC 1158
Db        301 ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro***IlePheGln*** 320

Qy       1159 AGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG 1218
Db        321 SerMetThrLysIleLeuGluProPheArg*********Pro***IleValIleTyrGln 340

Qy       1219 ------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATC 1272
Db        341 TyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIle 360

Qy       1273 GAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAG 1332
Db        361 GluGluLeuArg***HisLeuLeu***TrpGlyPhe***ThrProAspLysLysHisGln 380

Qy       1333 AAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
Db        381 LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGln 400

Qy       1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
Db        401 Pro******LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 420

Qy       1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
Db        421 LysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys***LysGlnLeuCys***Leu 440

Qy       1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
Db        441 LeuArgGlyAlaLysAlaLeuThrAspIleVal***LeuThrGluGluAlaGluLeuGlu 460

Qy       1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
Db        461 LeuAlaGluAsnArgGluIleLeuLys***ProValHisGlyValTyrTyrAspPro*** 480

Qy       1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
Db        481 LysAspLeuValAlaGlu***GlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyr 500

Qy       1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
Db        501 GlnGluProPheLysAsnLeuLysThrGlyLysTyrAla***LysArgSerAlaHisThr 520

Qy       1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
Db        521 AsnAspValLysGlnLeuThrGluValValGlnLysVal******GluSerIleValIle 540

Qy       1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
Db        541 TrpGlyLys***ProLysPheArgLeuProIleGln***GluThrTrpGlu***TrpTrp 560

Qy       1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
Db        561 MetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu 580

Qy       1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
Db        581 ValLysLeuTrpTyrGlnLeuGluLysAspProIle***GlyAlaGluThrPheTyrVal 600

Qy       1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
Db        601 AspGlyAlaAlaAsnArgGluThrLys***GlyLysAlaGlyTyrValThrAspArgGly 620

Qy       2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
Db        621 ArgGlnLysValValSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIle 640

Qy       2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
Db        641 HisLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAla 660

Qy       2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
Db        661 LeuGlyIleIleGlnAlaGlnProAspArgSerGluSerGlu***ValAsnGlnIleIle 680

Qy       2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
Db        681 GluLysLeuIle***Lys***LysValTyrLeuSerTrpValProAlaHisLysGlyIle 700

Qy       2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
Db        701 GlyGlyAsnGluGlnValAspLysLeuValSer***GlyIleArgLysValLeuPheLeu 720

Qy       2347 GACGGCATCGAT 2358
Db        721 AspGlyIleAsp 724
```

RESULT 14
US-10-301-661A-6
; Sequence 6, Application US/10301661A
; Publication No. US20030157660A1
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-
; APPLICANT: INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAUCLERE, Philippe
; APPLICANT: LOUSSERT-AJAKA, Ibtissam
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSSI, Francoise
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/10/301,661A
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US/09/319,588C
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-301-661A-6

Alignment Scores:

Pred. No.:	7.22e-145	Length:	1014
Score:	3140.50	Matches:	579
Percent Similarity:	89.70%	Conservative:	74
Best Local Similarity:	79.53%	Mismatches:	60
Query Match:	68.95%	Indels:	15
DB:	14	Gaps:	5

US-09-610-313B-32 (1-2457) x US-10-301-661A-6 (1-1014)

```
Qy        220 TTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db          1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProProAspAsn 20

Qy        280 AAC------CGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG--------------- 318
Db         21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40

Qy        319 ---CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG---------GGCACC 366
Db         41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db         61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80
```

```
Qy      427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
            :::::::||||||||||||||||||||||||||||||||||||||:::||||||:::
Db       81 GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluGluLeuGln 100

Qy      487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
            |||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      101 LeuGluGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArg 120

Qy      547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
            ||||||||   |||   ::::::|||   |||:::|||||:::|||||||||||:::
Db      121 GlnTyrAspAsnIleThrValAspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140

Qy      607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
            ||||||||||||||||||||||||||||||:::|||||||||:::|||||||||||
Db      141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160

Qy      667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180

Qy      727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
            ||||||||||||||||||||||||||||||||   ||||||||:::||||||   ||||||
Db      181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200

Qy      787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
               |||||||||||||||||||||||::::::||||||||||||||||||||||||
Db      201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220

Qy      847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
            |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      221 ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240

Qy      907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260

Qy      967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
            |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db      261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280

Qy     1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
            |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db      281 ProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGlu 300

Qy     1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db      301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320

Qy     1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
            ||||||||||:::|||||||||||||||||||||||||||   ::::::||||||||
Db      321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340

Qy     1207 GTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
            :::|||||||||      ||||||||||||||||||||||||:::   ||||||
Db      341 IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360

Qy     1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
            |||   :::|||:::||||||   |||||||||:::||||||||||||||||||||||
Db      361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380

Qy     1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGACAAG 1374
            ||||||||||||||||||||||||||||||                ||||||||||||||||
Db      381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400

Qy     1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
            ||||||||||||||||:::||||||||||||:::   ||||||||||||||||||||
Db      401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420

Qy     1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
            ||||||||||||||||||||||||||||||||||||||||||||||:::|||:::|||
Db      421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440

Qy     1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
            ||||||||||||:::|||||||||:::||||||||||::::::|||      |||||||||
Db      441 LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460

Qy     1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
            |||||||||||||||||||||||||||||||||||:::||||||:::|||||||||||
Db      461 AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480

Qy     1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
            ||||||||   |||:::|||||||||||||||||||||||||||||   |||||||||
Db      481 TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThr 500

Qy     1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
            ||||||||||||||||      ||||||||||||||||||||||||||||||||||||||
Db      501 TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520

Qy     1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
            :::|||||||||||||:::|||||||||   |||   |||:::|||:::|||   |||
Db      521 SerAlaHisThrAsnAspIleLysGlnLeuValGluValValArgLysValAlaThrGlu 540

Qy     1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
            |||||||||||||||||||||||||||||||||||||||:::|||||||||   |||
Db      541 SerIleValIleTrpGlyLysThrProLysPheArgLeuProValGlnLysGluValTrp 560

Qy     1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
            |||   ||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db      561 GluAlaTrpTrpThrAspHisTrpGlnAlaThrTrpIleProGluTrpGluPheValAsn 580

Qy     1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
            |||||||||||||||||||||||||||||||||   |||||||||   |||||||||
Db      581 ThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGluProIleSerGlyAlaGlu 600

Qy     1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
            ||||||||||||||||||||||||||||||||||||:::|||||||||||||:::|||
Db      601 ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620

Qy     2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
            |||||||||||||||||||:::||||||:::   :::||||||||||||||   |||
Db      621 ThrAspArgGlyArgGlnLysValValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640

Qy     2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
            |||||||||   :::|||||||||:::||||||   :::||||||||||||||||||
Db      641 LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660

Qy     2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
            ||||||||||||:::|||||||||   :::||||||||||||||||||||||||||||||
Db      661 SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680

Qy     2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
            :::||||||||||:::|||||||||||||||:::||||||||||||||||||||||||
Db      681 SerGlnIleIleGluGluLeuIleLysLysGluArgValTyrLeuSerTrpValProAla 700

Qy     2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
            |||||||||||||||||||||||||:::||||||||||||||   ||||||||||||
Db      701 HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720

Qy     2335 GTGCTGTTCCTGGACGGCATCGAT 2358
            :::|||||||||||||||||:::
Db      721 IleLeuPheLeuAspGlyIleGlu 728
```

RESULT 15
US-09-952-060-35
; Sequence 35, Application US/09952060
; Publication No. US20030044421A1
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.
; APPLICANT: Youil, Rima
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.

```
;  TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
;  TITLE OF INVENTION:  VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
;  TITLE OF INVENTION:  MODIFICATIONS
;  FILE REFERENCE: 20747Y
;  CURRENT APPLICATION NUMBER: US/09/952,060
;  CURRENT FILING DATE:  2001-09-14
;  PRIOR APPLICATION NUMBER: PCT/US01/28861
;  PRIOR FILING DATE: 2001-09-14
;  PRIOR APPLICATION NUMBER: 60/317,814
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: 60/279,056
;  PRIOR FILING DATE: 2001-03-27
;  PRIOR APPLICATION NUMBER: 60/233,180
;  PRIOR FILING DATE: 2000-09-15
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 1350
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Codon optimized gag-IA pol fusion
US-09-952-060-35

Alignment Scores:
Pred. No.:               9.56e-142      Length:        1350
Score:                   3076.50        Matches:       595
Percent Similarity:      80.66%         Conservative:  43
Best Local Similarity:   75.22%         Mismatches:    61
Query Match:             67.54%         Indels:        92
DB:                      10             Gaps:          7

US-09-610-313B-32 (1-2457) x US-09-952-060-35 (1-1350)

Qy         13 ATGGCCGAGGCCATGAGCCAGGCCACC---AGCGCCAACATCCTGATGCAGCGCAGCAAC 69
              :::|||||||||||||||   |||   ||||||   |||:::|||||||||   |||
Db        363 LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn 382

Qy         70 TTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCC 129
              |||:::      ::::::   :::||||||||||||||||||||   |||||||||||
Db        383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysValGlyHisIleAla 402

Qy        130 CGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAG 189
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422

Qy        190 ATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAG 249
              ||||||||||| |||||||||||||||||||   :::   |||||||||   ::
Db        423 MetLysAspCysAsnGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProSerHisLy 442

Qy        250 GGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAG 309
              :  ||||||   :::         ||||||     ||||||||||||||      ||
Db        442 sGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProGluGluSe 462

Qy        310 CTGCAGGTGCGCG------GCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGC 363
              |   |||              |||||||||:::   :::       |||       :::
Db        462 rPheArgPheGlyGluGluLysThrThrProSerGlnLysGlnGluProIleAspLys-- 481

Qy        364 ACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGC 423
                  |||                              ||||||   |||::::::
Db        482 GluLeu---------------------------TyrProLeuAlaSerLeuArg------ 490

Qy        424 GGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATG 483

Db        490 ------------------------------------------------------------ 490

Qy        484 AGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTG 543

Db        490 ------------------------------------------------------------ 490

Qy        544 CGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTG 603

Db        490 ------------------------------------------------------------ 490

Qy        604 ATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACC 663
                                 ::::::   |||   :::      :::|||:::
Db        491 ------------------SerLeuPheGlyAsnAspProSerSerGlnMetAla------ 502

Qy        664 CTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGAC 723
                       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 ---------ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAsp 519

Qy        724 GGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATC 783
              ||||||||||||||||||||||||||||||||||||||||||||||||||     |||
Db        520 GlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIle 539

Qy        784 TGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAAC 843
              |||   ||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        540 CysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsn 559

Qy        844 ACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTC 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPhe 579

Qy        904 CGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCC 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPro 599

Qy        964 GCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGC 1023
              |||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        600 AlaGlyLeuLysLysLysLysSerValThrValLeuAlaValGlyAspAlaTyrPheSer 619

Qy       1024 GTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAAC 1083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 ValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsn 639

Qy       1084 GAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCC 1143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 GluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro 659

Qy       1144 AGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAG 1203
              :::|||||||||||||||||||||||||||||||||||||||   :::||||||:::
Db        660 AlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAsp 679

Qy       1204 ATCGTGATCTACCAG------GCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAG 1257
              |||||||||||||||      |||   |||||||||||||||||||||||||||||||
Db        680 IleValIleTyrGlnTyrMetAlaAlaLeuTyrValGlySerAspLeuGluIleGlyGln 699

Qy       1258 CACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCC 1317
              ||||||   ||||||||||||||||||:::|||||||||||||||||||   |||||||||
Db        700 HisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrPro 719

Qy       1318 GACAAGAAGCACCAGAAGGAGCCCCCCTTCCTG------CCCATCGAGCTGCACCCCGAC 1371
              |||||||||||||||||||||||||||||||||            |||||||||||||||
Db        720 AspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAsp 739

Qy       1372 AAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATC 1431
              |||||||||||||||||||   ||||||||||||:::||||||||||||||||||||||
Db        740 LysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIle 759

Qy       1432 CAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGC 1491
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArg 779

Qy       1492 CAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAG 1551
              ||||||||||||||||||||||||   ||||||||||||:::::::::||||||||||||
Db        780 GlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGlu 799

Qy       1552 GAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTG 1611
              |||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        800 GluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyVal 819

Qy       1612 TACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGG 1671
              ||||||||||||||||||||||||:::||||||||||||||||||||||     ||||||
Db        820 TyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrp 839
```

```
Qy       1672 ACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATG 1731
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        840 ThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859

Qy       1732 CGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATG 1791
              |||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        860 ArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr 879

Qy       1792 GAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACC 1851
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        880 GluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGluThr 899

Qy       1852 TGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTG 1911
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        900 TrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVal 919

Qy       1912 AACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCC 1971
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        920 AsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAla 939

Qy       1972 GAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTAC 2031
              |||||||||||||||      ||||||||||||||||||:::||||||||||||||||||
Db        940 GluThrPheTyrValAlaGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyr 959

Qy       2032 GTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACC 2091
              ||||||:::|||||||||||||||:::|||:::||||||:::|||||||||||||||||||||
Db        960 ValThrAsnArgGlyArgGlnLysValValThrLeuThrAspThrThrAsnGlnLysThr 979

Qy       2092 GAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACC 2151
                 ||||||||||||   |||||||||||||||||||||   ||||||||||||||||||
Db        980 AlaLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThr 999

Qy       2152 GACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTG 2211
                 |||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db       1000 AlaSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeu 1019

Qy       2212 GTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCC 2271
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       1020 ValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValPro 1039

Qy       2272 GCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGC 2331
              ||||||||||||||||||||||||||||||:::|||||||||||||||   |||||||||
Db       1040 AlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArg 1059

Qy       2332 AAGGTGCTGTTCCTGGACGGCATCGAT 2358
              |||||||||||||||||||||||||||
Db       1060 LysValLeuPheLeuAspGlyIleAsp 1068
```

Search completed: June 2, 2005, 05:29:26
Job time : 221.794 secs